# BẮT ĐẦU VỚI NGÔN NGỮ C

## PHẦN 1.1 HELLO WORLD

Để tạo một chương trình C đơn giản in dòng chữ "Hello, World " lên màn hình, sử dụng trình soạn thảo văn bản để tạo file mới (ví dụ:

hello.c — phần mở rộng file phải là .c) chứa mã nguồn sau:

**hello.c**

|  |
| --- |
| # include<stdio.h>  int main(void)  {      puts("Hello, World");      return 0;  } |

**Hãy xem từng dòng trong chương trình đơn giản này**

|  |
| --- |
| # include<stdio.h> |

Dòng mã này yêu cầu trình biên dịch cần thêm nội dung của tệp tiêu đề thư viện chuẩn **stdio.h** vào chương trình.

Tệp tiêu đề thường chứa các khai báo của các hàm, macros và kiểu dữ liệu trong thư viện chuẩn.

và bạn cần phải thêm vào tệp tiêu đề trước khi sử dụng chúng. Dòng này thêm vào thư viện **stdio.h** để nó có thể gọi hàm **puts().**

Tìm hiểu thêm về tệp tiêu đề

|  |
| --- |
| int main(void) |

Dòng mã này bắt đầu định nghĩa của một hàm. Nó cho biết tên của hàm (main), kiểu dữ liệu và số lượng các đối số mà nó mong đợi (void, nghĩa là không có) và loại giá trị mà hàm này trả về (int). Theo nguyên tắc, **main** là một tên đặc biệt, nó chỉ dùng để đặt cho function chính của chương trình, và lúc nào chương trình cũng sẽ bắt đầu từ **function main**.

Việc thực thi chương trình bắt đầu trong hàm main().

|  |
| --- |
| {      ...  } |

Dấu ngoặc nhọn (curly braces) được sử dụng thành cặp để chỉ định vị trí bắt đầu và kết thúc một khối mã(khối lệnh). Chúng có thể được sử dụng theo nhiều cách khác nhau, nhưng trong trường hợp này, chúng chỉ định vị trí bắt đầu và kết thúc của một hàm.

|  |
| --- |
| puts("Hello, World"); |

Dòng này gọi hàm puts() để đưa ra văn bản ra đầu ra tiêu chuẩn (mặc định là màn hình), tiếp theo là một dòng mới.

Chuỗi cần đưa ra được bao gồm trong cặp dấu ngoặc đơn().

"Hello, World" là chuỗi sẽ được ghi ra màn hình. Trong ngôn ngữ C, mọi giá trị chuỗi phải nằm trong dấu ngoặc kép "...".

Tìm hiểu thêm về chuỗi.

Trong các chương trình C, mỗi câu lệnh cần được kết thúc bằng dấu chấm phẩy (tức là **;**).

|  |
| --- |
| return 0; |

Khi chúng ta định nghĩa hàm **main(),** chúng ta khai báo nó là một hàm trả về kiểu **int**, có nghĩa là nó cần trả về một giá trị số nguyên. Trong ví dụ này, chúng ta đang trả về giá trị nguyên là **0**, được sử dụng để chỉ rằng chương trình đã kết thúc thành công. Sau câu lệnh **return 0;**, quá trình thực thi sẽ kết thúc.

**Chỉnh sửa chương trình**

Các trình soạn thảo, chỉnh sửa văn bản đơn giản bao gồm **vim** hoặc **gedit** trên **Linux,** hoặc **Notepad** trên **Windows.** Các trình soạn thảo đa nền tảng khác bao gồm **Visual Studio Code** hoặc **Sublime Text.**

Trình chỉnh sửa phải tạo các tệp văn bản thuần túy, không phải RTF hoặc bất kỳ định dạng nào khác.

**Biên dịch và chạy chương trình**

Để chạy chương trình, file nguồn này (hello.c) trước tiên cần được biên dịch thành file thực thi (ví dụ: **hello** trên hệ thống **Unix/Linux** hoặc **hello.exe** trên **Windows**). Điều này được thực hiện bằng cách sử dụng trình biên dịch cho ngôn ngữ C.

**Biên dịch bằng GCC**

GCC **(GNU Compiler Collection)** là một trình biên dịch C được sử dụng rộng rãi. Để sử dụng nó, hãy mở một thiết bị đầu cuối, sử dụng dòng lệnh để điều hướng đến vị trí của tệp nguồn rồi chạy:

|  |
| --- |
| gcc hello.c -o hello |

Nếu không tìm thấy lỗi trong mã nguồn (hello.c), trình biên dịch sẽ tạo một tệp nhị phân, tên của tệp được cung cấp bởi đối số cho tùy chọn dòng lệnh -o (hello). Đây là tập tin thực thi cuối cùng.

Chúng tôi cũng có thể sử dụng các tùy chọn cảnh báo *-Wall -Wextra -Werror*, giúp xác định các sự cố có thể khiến chương trình bị lỗi hoặc tạo ra kết quả không mong muốn. Chúng không cần thiết cho chương trình đơn giản này nhưng đây là cách thêm chúng:

|  |
| --- |
| gcc -Wall -Wextra -Werror -o hello hello.c |

**Sử dụng trình biên dịch clang**

Để biên dịch chương trình bằng **clang**, bạn có thể sử dụng:

|  |
| --- |
| clang -Wall -Wextra -Werror -o hello hello.c |

Theo thiết kế, các tùy chọn dòng lệnh **clang** tương tự như của **GCC.**

**Sử dụng trình biên dịch Microsoft C từ dòng lệnh**

Nếu sử dụng trình biên dịch Microsoft cl.exe trên hệ thống Windows hỗ trợ Visual Studio và nếu tất cả các biến môi trường được đặt, ví dụ C này có thể được biên dịch bằng lệnh sau. Lệnh này sẽ tạo ra tệp hello.exe có thể thực thi được trong thư mục mà lệnh được thực thi trong (Có các tùy chọn cảnh báo như /W3 cho cl, đại khái là tương tự như -Wall vv cho GCC hoặc clang).

|  |
| --- |
| cl hello.c |

**Thực hiện chương trình**

Sau khi được biên dịch, tệp nhị phân sau đó có thể được thực thi bằng cách nhập ./hello trong thiết bị đầu cuối. Khi thực thi, chương trình đã biên dịch sẽ in **Hello, World**, theo sau là một dòng mới, tới dấu nhắc lệnh.

## PHẦN 1.2 BẢN GỐC “HELLO WORLD!” TRONG K&R C

Sau đây là bản gốc "Hello World!" chương trình từ cuốn sách **The C Programming Language** của BrianKernighan và Dennis Ritchie (Ritchie là nhà phát triển ban đầu của ngôn ngữ lập trình C tại Bell Labs), được gọi là "K&R":

|  |
| --- |
| Version = K&R  #include <stdio.h>  main()  {   printf("hello, world\n");  } |

Lưu ý rằng **The C Programming Language** không được chuẩn hóa vào thời điểm viết ấn bản đầu tiên của cuốn sách này (1978), và chương trình này có thể sẽ không được biên dịch trên hầu hết các trình biên dịch hiện đại trừ khi chúng được hướng dẫn chấp nhận mã C90.

Câu ví dụ đầu tiên trong cuốn sách K&R (**The C Programming Language**) hiện được coi là chất lượng kém, một phần là do thiếu một kiểu trả về rõ ràng cho **main()** và một phần là do thiếu câu lệnh **return**. Phiên bản thứ hai của cuốn sách được viết cho tiêu chuẩn C89 cũ. Trong C89, kiểu của main sẽ mặc định là int, nhưng ví dụ K&R không trả về một giá trị xác định cho môi trường. Trong các tiêu chuẩn C99 và sau này, kiểu trả về là bắt buộc, nhưng có thể bỏ qua câu lệnh **return** của main (**và chỉ main**), vì có một trường hợp đặc biệt được giới thiệu với C99 5.1.2.2.3 - nó tương đương với việc trả về 0, biểu thị thành công.

Dạng chính được khuyến nghị và di động nhất cho các hệ thống được lưu trữ là **int main(void)** khi chương trình không sử dụng bất kỳ đối số dòng lệnh nào, hoặc **int main(int argc, char \*\*argv)** khi chương trình sử dụng các đối số ở các dòng lệnh.

**C90 §5.1.2.2.3 Kết thúc chương trình**

Một câu lệnh **return** từ lời gọi ban đầu tới hàm **main** tương đương với việc gọi hàm **exit** với giá trị trả về của hàm **main** là **đối số**. Nếu hàm main thực hiện một câu lệnh **return** mà không xác định giá trị, trạng thái kết thúc trả về cho môi trường chủ không được xác định.

**C90 §6.6.6.4 Câu lệnh return**

Nếu một câu lệnh **return** mà không có biểu thức được thực thi và giá trị của lời gọi hàm được sử dụng bởi người gọi, hành vi đó là không xác định. Đạt đến dấu **}** kết thúc một hàm tương đương với việc thực thi một câu lệnh **return** mà không có biểu thức.

**C99 §5.1.2.2.3 Kết thúc chương trình**

Nếu kiểu trả về của hàm **main** là một kiểu tương thích với **int**, một câu lệnh **return** từ lời gọi ban đầu tới hàm **main** tương đương với việc gọi hàm **exit** với giá trị trả về của hàm main là đối số; đạt đến dấu **}** kết thúc hàm main sẽ trả về giá trị 0. Nếu kiểu trả về không tương thích với int, trạng thái kết thúc trả về cho môi trường chủ không được xác định.

# COMMENTS

Chú thích (comments) được sử dụng để chỉ điều gì đó cho người đọc mã nguồn. Chú thích được coi như một phần trống trên mã nguồn và không thay đổi bất kỳ ý nghĩa thực sự nào của mã. Có hai cú pháp được sử dụng cho chú thích trong ngôn ngữ C, đó là **/\* \*/** (bản gốc) và **//** (bản mới hơn 1 chút). Một số hệ thống tài liệu sử dụng chú thích được định dạng đặc biệt để hỗ trợ tạo tài liệu cho mã nguồn.

## PHẦN 2.1 CHÚ THÍCH BẰNG CÁCH SỬ DỤNG TIỀN XỬ LÝ(PREPROCESSOR)

Một phần lớn mã nguồn cũng có thể bị "chú thích" bằng cách sử dụng các chỉ thị tiền xử lý **#if 0** và **#endif.** Điều này hữu ích khi mã chứa các chú thích nhiều dòng mà không thể lồng nhau nếu không có cơ chế chú thích này.

|  |
| --- |
| #if 0 /\* Bắt đầu "chú thích", bất cứ điều gì từ đây trở đi sẽ bị loại bỏ bởi tiền xử lý \*/  /\* Một lượng mã lớn với các chú thích nhiều dòng \*/  int foo()  {  /\* một số câu lệnh \*/  ...  /\* ... một chú thích mô tả câu lệnh if ... \*/  if (someTest) {   /\* some more comments \*/  return 1;  }  return 0;  }  #endif /\* 0 \*/  /\* Mã từ đây trở đi không bị "chú thích" (được bao gồm trong tệp thực thi biên dịch) \*/ |

## PHẦN 2.2 CHÚ THÍCH ĐƯỢC GIỚI HẠN BẰNG /\* \*/

Một chú thích bắt đầu với một dấu gạch chéo kề ngay sau đó là một dấu hoa thị (/\*) và kết thúc ngay khi gặp một dấu hoa thị kề ngay sau đó là một dấu gạch chéo (\*/). Mọi thứ nằm giữa các cặp ký tự này đều là chú thích và được coi là một phần trống (tức là bị bỏ qua) bởi trình biên dịch.

|  |
| --- |
| /\* Đây là một chú thích \*/ |

Chú thích ở trên là chú thích trên một dòng. Chú thích kiểu /\* này có thể trải dài trên nhiều dòng, như sau:

|  |
| --- |
| /\* Đây là một chú thích  nhiều dòng \*/ |

Mặc dù điều này không bắt buộc, quy ước phong cách thông thường với chú thích trên nhiều dòng là đặt khoảng trắng và dấu hoa thị ở các dòng sau đầu tiên và đặt /\* và \*/ trên các dòng mới, sao cho chúng xếp thành hàng:

|  |
| --- |
| /\* Đây là   \*một chú   \* thích   \*nhiều dòng  \*/ |

Những dấu hoa thị thêm không có tác dụng chức năng nào đối với chú thích vì chúng không có dấu gạch chéo liên quan.

Các chú thích kiểu /\* này có thể được sử dụng trên một dòng riêng biệt, ở cuối một dòng mã hoặc ngay cả trong các dòng mã:

|  |
| --- |
| /\* Chú thích này nằm trên một dòng riêng biệt \*/  if(x && y) { /\* Chú thích này nằm ở cuối dòng mã \*/      if ((complexCondition1) /\* Chú thích này nằm trong một dòng mã \*/          && (complexCondition2)) {      /\* Chú thích này nằm trong một câu lệnh if, trên một dòng riêng biệt \*/      }  } |

Chú thích không thể lồng nhau. Điều này xảy ra vì bất kỳ /\* nào tiếp theo sẽ bị bỏ qua (là một phần của chú thích) và đầu tiên \*/ gặp được sẽ được coi là kết thúc chú thích. Chú thích trong ví dụ dưới đây sẽ không hoạt động:

|  |
| --- |
| /\* Chú thích bên ngoài, có nghĩa là điều này bị bỏ qua => /\* chú thích nội bộ cố gắng \*/ <= kết thúc chú thích,  không phải chú thích này => \*/ |

Để chú thích các khối mã chứa chú thích kiểu này, mà nếu không sẽ bị lồng nhau, bạn có thể tham khảo ví dụ về Chú thích bằng cách sử dụng tiền xử lý.

## PHẦN 2.3 CHÚ THÍCH ĐƯỢC GIỚI HẠN BẰNG DẤU //

Phiên bản ≥ C99

C99 đã giới thiệu việc sử dụng chú thích trên một dòng theo kiểu C++. Loại chú thích này bắt đầu bằng hai dấu gạch chéo kề nhau và kéo dài đến cuối dòng:

|  |
| --- |
| // Mỗi dòng trong số này đều là một chú thích trên một dòng  // Chú ý rằng mỗi dòng phải bắt đầu bằng  // hai dấu gạch chéo kép liên tiếp (//) |

Loại chú thích này có thể được sử dụng trên một dòng riêng biệt hoặc ở cuối một dòng mã. Tuy nhiên, do chú thích này kéo dài đến cuối dòng, nên chúng không thể được sử dụng trong một dòng mã.

|  |
| --- |
| // Chú thích này nằm trên một dòng riêng biệt  if (x && y) { // Chú thích này nằm ở cuối dòng  // Chú thích này nằm trong một câu lệnh if, trên một dòng riêng biệt  } |

## PHẦN 2.4 RỦI RO CÓ THỂ XẢY RA DO TRIGRAPH

Trong quá trình viết chú thích được giới hạn bởi //, có thể xảy ra lỗi chính tả ảnh hưởng đến hoạt động dự kiến của chú thích. Nếu bạn gõ sai như sau:

|  |
| --- |
| int a = 10 //Sao tôi làm điều này??/ |

Dấu gạch chéo / ở cuối là một lỗi chính tả nhưng giờ đây sẽ được hiểu là dấu gạch chéo ngược . Điều này xảy ra vì ??/ tạo thành một trigraph. Trigraph ??/ thực tế là một cách viết dài cho dấu gạch chéo ngược , đây là ký hiệu tiếp tục dòng. Điều này có nghĩa là trình biên dịch sẽ hiểu rằng dòng tiếp theo là tiếp tục của dòng hiện tại, tức là tiếp tục của chú thích, điều này có thể không phải ý định ban đầu.

|  |
| --- |
| int foo = 20; // Bắt đầu từ 20 ??/  int bar = 0;  // Dòng sau sẽ gây lỗi biên dịch (biến 'bar' chưa được khai báo)  // vì 'int bar = 0;' là một phần của chú thích trên dòng trước đó  bar += foo; |

# KIỂU DỮ LIỆU

## PHẦN 3.1 GIẢI THÍCH CÁC KHAI BÁO

Một đặc điểm cú pháp đặc biệt của ngôn ngữ C là các khai báo phản ánh việc sử dụng của đối tượng được khai báo như trong một biểu thức thông thường.

Tập hợp các toán tử sau đây có cùng mức độ ưu tiên và tính kết hợp được tái sử dụng trong các khai báo, bao gồm:

* Toán tử unary \* "dereference" (giải tham chiếu) thể hiện một con trỏ.
* Toán tử binary [] "array subscription" (phần tử mảng) thể hiện một mảng.
* Toán tử (1+n)-ary () "function call" (gọi hàm) thể hiện một hàm.
* Dấu ngoặc () nhóm nhưng ghi đè mức độ ưu tiên và tính kết hợp của các toán tử còn lại được liệt kê.

Ba toán tử trên có độ ưu tiên và tính kết hợp như sau:

|  |  |  |
| --- | --- | --- |
| Toán tử | Mức độ ưu tiên | Tính kết hợp |
| [ ] Phần tử mảng | 1 | Trái sang phải |
| ( ) Gọi hàm | 1 | Trái sang phải |
| \*Giải tham chiếu | 2 | Phải sang trái |

Khi giải thích các khai báo, người đọc cần bắt đầu từ từng định danh và áp dụng các toán tử kề cạnh theo thứ tự đúng như trong bảng trên. Mỗi lần áp dụng một toán tử, ta có thể thay thế nó bằng các từ tiếng Anh sau đây:

|  |  |
| --- | --- |
| Biểu thức | Giải thích |
| Thing[X] | Kích thức của mảng X là… |
| Thing(x1,x2,x3) | 1 hàm nhận x1 x2 x3 và trả về … |
| \*Thing | Một con trỏ trỏ tới .. |

Theo đó phần bắt đầu của việc giải thích bằng tiếng Anh sẽ luôn bắt đầu với định danh và kết thúc với kiểu dữ liệu nằm ở phía bên trái của khai báo.

**Ví dụ**

|  |
| --- |
| char \*names[20]; |

[] được ưu tiên hơn \*, vì vậy cách hiểu là: **names** là một mảng có kích thước 20 của một con trỏ trỏ tới char.

|  |
| --- |
| char (\*place)[10]; |

Trong trường hợp sử dụng dấu ngoặc () để ghi đè mức độ ưu tiên, toán tử \* được áp dụng trước: **place** là một con trỏ tới một mảng có kích thước là 10 của char.

|  |
| --- |
| int fn(long, short); |

Ở đây không có ưu tiên cần quan tâm: fn là một hàm nhận tham số long và short, và trả về kiểu int.

|  |
| --- |
| int \*fn(void); |

Dấu ngoặc () được áp dụng trước: fn là một hàm nhận tham số void và trả về một con trỏ tới kiểu int.

|  |
| --- |
| int (\*fp)(void); |

Ghi đè mức độ ưu tiên của (): fp là một con trỏ tới một hàm nhận tham số void và trả về kiểu int.

|  |
| --- |
| int arr[5][8]; |

Mảng đa chiều không phải là một ngoại lệ trong quy tắc; các toán tử [] được áp dụng theo thứ tự từ trái sang phải theo tính kết hợp được liệt kê trong bảng: arr là một mảng có kích thước 5 của một mảng có kích thước 8 của kiểu int.

|  |
| --- |
| int \*\*ptr; |

Hai toán tử dereference có cùng mức độ ưu tiên, do đó tính kết hợp có hiệu lực. Các toán tử được áp dụng theo thứ tự từ phải sang trái: ptr là một con trỏ tới một con trỏ tới kiểu int.

**Khai báo nhiều biến**

Dấu phẩy có thể được sử dụng làm dấu phân tách (không hoạt động như toán tử dấu phẩy) để phân cách các khai báo đa biến trong cùng một câu lệnh. Câu lệnh sau chứa năm khai báo:

|  |
| --- |
| int fn(void), \*ptr, (\*fp)(int), arr[10][20], num; |

Các đối tượng được khai báo trong ví dụ trên là:

* fn: một hàm không tham số và trả về kiểu int.
* ptr: một con trỏ tới kiểu int.
* fp: một con trỏ hàm nhận tham số kiểu int và trả về kiểu int.
* arr: một mảng có kích thước 10, mỗi phần tử là một mảng có kích thước 20 kiểu int.
* num: kiểu int.

**Giải thích thay thế**

Bởi vì các khai báo phản ánh việc sử dụng, nên một khai báo cũng có thể được giải thích dựa trên các toán tử có thể được áp dụng lên đối tượng và kiểu dữ liệu cuối cùng của biểu thức đó. Kiểu dữ liệu nằm ở phía bên trái là kết quả cuối cùng sau khi áp dụng tất cả các toán tử.

|  |
| --- |
| /\*   \*Sử dụng toán tử [] để lấy phần tử của "arr" và sau đó dùng toán tử dereference để lấy giá trị,   \*ta được một kết quả kiểu "char".   \*Cụ thể: \*arr[5] có kiểu "char".   \*/  char arr[20];  /\*   \*Gọi hàm "fn" trả về một kết quả kiểu "int".   \*Cụ thể: fn('b') có kiểu "int".   \*/  int fn(char);  /\*   \*Sử dụng toán tử dereference để lấy giá trị của "fp", sau đó gọi nó trả về một kết quả kiểu "int".   \*Cụ thể: (\*fp)() có kiểu "int".   \*/  int (fp)(void);  /\*   \*Sử dụng toán tử [] hai lần để lấy phần tử của "strings", sau đó sử dụng toán tử dereference để lấy giá trị,   \*ta được một kết quả kiểu "char".   \*Cụ thể: \*strings[5][15] có kiểu "char".   \*/  char \*strings[10][20]; |

## PHẦN 3.2 KIỂU SỐ NGUYÊN CÓ ĐỘ RỘNG CỐ ĐỊNH ( KỂ TỪ C99)

Phiên bản ≥ C99 Tiêu đề <stdint.h> cung cấp một số định nghĩa kiểu số nguyên có độ rộng cố định. Các kiểu này là tùy chọn và chỉ được cung cấp nếu nền tảng có một kiểu số nguyên có độ rộng tương ứng và nếu kiểu có dấu tương ứng có biểu diễn bù hai của các giá trị âm. Xem phần lưu ý để biết các gợi ý về việc sử dụng các kiểu có độ rộng cố định.

|  |
| --- |
| /\* Các kiểu phổ biến thường được sử dụng bao gồm \*/  *uint32\_t* u32 = 32; / có độ rộng chính xác 32 bit /  uint8\_t u8 = 255; / có độ rộng chính xác 8 bit /  int64\_t i64 = -65; / có độ rộng chính xác 64 bit trong biểu diễn bù hai \*/ |

## PHẦN 3.3 KIỂU SỐ NGUYÊN VÀ HẰNG SỐ

Các số nguyên có dấu có thể thuộc các kiểu sau đây (int sau từ short hoặc long là tùy chọn):

|  |
| --- |
| signed char c = 127; /\* yêu cầu có độ rộng 1 byte, xem phần lưu ý để biết thêm thông tin \*/  signed short int si = 32767; /\* yêu cầu có ít nhất 16 bit \*/  signed int i = 32767; /\* yêu cầu có ít nhất 16 bit \*/  signed long int li = 2147483647; /\* yêu cầu có ít nhất 32 bit \*/  Phiên bản ≥ C99  signed long long int li = 2147483647; /\* yêu cầu có ít nhất 64 bit \*/ |

Mỗi kiểu số nguyên có dấu này đều có một phiên bản không dấu tương ứng.

|  |
| --- |
| unsigned int i = 65535;  unsigned short = 2767;  unsigned char = 255; |

Đối với tất cả các kiểu, ngoại trừ char, phiên bản có dấu được cho là mặc định nếu phần có dấu hoặc không dấu bị bỏ qua. Kiểu char tạo thành một kiểu ký tự thứ ba, khác với signed char và unsigned char, và sự có dấu (hoặc không) phụ thuộc vào nền tảng.

Có các loại hằng số số nguyên khác nhau (được gọi là chữ số trong ngôn ngữ C) có thể được viết theo các cơ số khác nhau và có độ rộng khác nhau, dựa trên tiền tố hoặc hậu tố của chúng.

|  |
| --- |
| /\* Các biến sau được khởi tạo với cùng một giá trị: \*/  int d = 42; /\* hằng số thập phân (cơ số 10) \*/  int o = 052; /\* hằng số bát phân (cơ số 8) \*/  int x = 0xaf; /\* hằng số thập lục phân (cơ số 16) \*/  int X = 0XAf; /\*\*/ (chữ cái 'a' đến 'f' (không phân biệt chữ hoa chữ thường) biểu thị từ 10 đến 15) \*/ |

Hằng số thập phân luôn luôn có dấu**(signed)**. Hằng số thập lục phân bắt đầu bằng 0x hoặc 0X và hằng số bát phân bắt đầu chỉ với chữ số 0. Hai loại hằng số sau có thể có dấu**(signed)** hoặc không dấu**(unsigned)** tùy thuộc vào việc giá trị có phù hợp với kiểu có dấu hay không.

|  |
| --- |
| /\* Tiền tố để chỉ độ rộng và dấu \*/  long int i = 0x32; /\* không có tiền tố đại diện cho int hoặc long int \*/  unsigned int ui = 65535u; /\* u hoặc U đại diện cho unsigned int hoặc unsigned long int \*/  long int li = 65536l; /\* l hoặc L đại diện cho long int \*/ |

Nếu không có tiền tố, hằng số có kiểu dữ liệu là kiểu đầu tiên mà giá trị của nó phù hợp, nghĩa là hằng số thập phân lớn hơn INT\_MAX sẽ có kiểu long nếu có thể, hoặc long long nếu không được.

Tập tin tiêu đề <limits.h> mô tả các giới hạn của số nguyên như sau. Các giá trị xác định bởi người thực hiện (implementation-defined) phải có giá trị tuyệt đối lớn hơn hoặc bằng giá trị được hiển thị dưới đây, với cùng dấu.

|  |  |  |
| --- | --- | --- |
| MARCO | Kiểu dữ liệu | Giá trị |
| CHAR\_BIT | smallest object that is not a bit-field (byte) 8 |  |
| SCHAR\_MIN | signed char | -127 / -(27 - 1) |
| SCHAR\_MAX | signed char | +127 / 27 - 1 |
| UCHAR\_MAX | unsigned char | 255 / 28 - 1 |
| CHAR\_MIN | char | see below |
| CHAR\_MAX | char | see below |
| SHRT\_MIN | short int | -32767 / -(215 - 1) |
| SHRT\_MAX | short int | +32767 / 215 - 1 |
| USHRT\_MAX | unsigned short int | 65535 / 216 - 1 |
| INT\_MIN | int | -32767 / -(215 - 1) |
| INT\_MAX | int | +32767 / 215 - 1 |
| UINT\_MAX | unsigned int | 65535 / 216 - 1 |
| LONG\_MIN | long int | -2147483647 / -(231 - 1) |
| LONG\_MAX | long int | +2147483647 / 231 - 1 |
| ULONG\_MAX | unsigned long int | 4294967295 / 232 - 1 |
| Version ≥ C99 |  |  |
| LLONG\_MIN | long long int | -9223372036854775807 / -(263 - 1) |
| LLONG\_MAX | long long int | +9223372036854775807 / 263 - 1 |
| ULLONG\_MAX | unsigned long long int | 18446744073709551615 / 264 - 1 |

Nếu giá trị của một đối tượng có kiểu char được mở rộng dấu khi sử dụng trong một biểu thức, giá trị của CHAR\_MIN sẽ giống như SCHAR\_MIN và giá trị của CHAR\_MAX sẽ giống như SCHAR\_MAX. Nếu giá trị của một đối tượng có kiểu char không được mở rộng dấu khi sử dụng trong một biểu thức, giá trị của CHAR\_MIN sẽ là 0 và giá trị của CHAR\_MAX sẽ giống như UCHAR\_MAX.

Phiên bản ≥ C99

Tiêu chuẩn C99 đã thêm một tiêu đề mới, <stdint.h>, chứa các định nghĩa cho các số nguyên có chiều rộng cố định. Xem ví dụ số nguyên chiều rộng cố định để được giải thích sâu hơn.

## PHẦN 3.4 HẰNG SỐ THỰC

Ngôn ngữ C có ba kiểu số thực dấu phẩy động bắt buộc, đó là **float**, **double** và **long double**.

|  |
| --- |
| float f = 0.314f; /\* Hậu tố f hoặc F chỉ định kiểu float \*/  double d = 0.314; /\* Không có hậu tố chỉ định kiểu double \*/  long double ld = 0.314l; /\* Hậu tố l hoặc L chỉ định kiểu long double \*/  /\* Các phần khác nhau trong một định nghĩa số thực là tùy chọn \*/  double x = 1.; /\* Hợp lệ, phần thập phân là tùy chọn \*/  double y = .1; /\* Hợp lệ, phần nguyên là tùy chọn \*/  /\* Chúng cũng có thể được định nghĩa trong dạng ký hiệu khoa học \*/  double sd = 1.2e3; /\* Phần thập phân 1.2 được nhân với 10^3, tức là 1200.0 \*/ |

Tiêu đề <float.h> xác định các giới hạn khác nhau cho các hoạt động của dấu phẩy động Số học dấu phẩy động được xác định triển khai. Tuy nhiên, hầu hết các nền tảng hiện đại (arm, x86, x86\_64, MIPS) đều sử dụng các hoạt động của dấu phẩy động IEEE 754. C cũng có ba loại dấu phẩy động phức tạp tùy chọn bắt nguồn từ loại trên.

## PHẦN 3.5 CHUỖI KÍ TỰ

Một chuỗi ký tự trong ngôn ngữ C là một chuỗi các ký tự, kết thúc bằng một ký tự số không (literal zero)

|  |
| --- |
| char\* str = "hello, world"; /\* chuỗi ký tự nguyên mẫu (string literal) \*/  /\* chuỗi ký tự nguyên mẫu có thể được sử dụng để khởi tạo mảng \*/  char a1[] = "abc"; /\* a1 là mảng char[4] chứa {'a','b','c','\0'} \*/  char a2[4] = "abc"; /\* tương tự a1 \*/  char a3[3] = "abc"; /\* a3 là mảng char[3] chứa {'a','b','c'}, thiếu ký tự '\0' \*/ |

Chuỗi ký tự nguyên mẫu trong C **không thể sửa đổi**, và việc cố gắng thay đổi giá trị của chúng sẽ dẫn đến hành vi không xác định. Chuỗi ký tự nguyên mẫu thường được lưu trữ trong bộ nhớ chỉ đọc (read-only memory), chẳng hạn như phần .rodata, và mọi cố gắng sửa đổi giá trị của chúng có thể dẫn đến kết quả không mong muốn.

|  |
| --- |
| char\* s = "foobar";  s[0] = 'F'; /\* hành vi không xác định \*/  /\* Thực hành tốt là đánh dấu chuỗi nguyên mẫu là hằng số bằng cách sử dụng const \*/  char const s1 = "foobar";  s1[0] = 'F'; /\* lỗi biên dịch! \*/ |

Nhiều chuỗi ký tự được nối vào thời gian biên dịch, có nghĩa là bạn có thể viết cấu trúc như thế này.

|  |
| --- |
| Phiên bản trước C99  /\* Chỉ có thể nối hai chuỗi ký tự hẹp (narrow) hoặc hai chuỗi ký tự rộng (wide) \*/  char\* s = "Hello, " "World";  Phiên bản C99 trở lên  /\* Từ phiên bản C99 trở đi, có thể nối nhiều hơn hai chuỗi \*/  /\* Quá trình nối chuỗi phụ thuộc vào triển khai cụ thể \*/  char\* s1 = "Hello" ", " "World";  /\* Một số cách sử dụng phổ biến là nối chuỗi định dạng (format strings) \*/  char\* fmt = "%" PRId16; /\* Macro PRId16 được định nghĩa từ C99 \*/ |

Chuỗi ký tự nguyên mẫu, tương tự như hằng số ký tự, hỗ trợ các bộ ký tự khác nhau.

|  |
| --- |
| /\* Chuỗi ký tự thông thường, kiểu char[] \*/  char\* s1 = "abc";  /\* Chuỗi ký tự rộng (wide character), kiểu wchar\_t[] \*/  *wchar\_t*\* s2 = L"abc";  Phiên bản C11 trở lên  /\* Chuỗi ký tự UTF-8, kiểu char[] \*/  char\* s3 = u8"abc";  /\* Chuỗi ký tự rộng 16-bit, kiểu char16\_t[] \*/  *char16\_t*\* s4 = u"abc";  /\* Chuỗi ký tự rộng 32-bit, kiểu char32\_t[] \*/  *char32\_t*\* s5 = U"abc"; |

# TOÁN TỬ

Toán tử trong ngôn ngữ lập trình là một ký hiệu cho biết trình biên dịch hoặc trình thông dịch thực hiện một phép toán, phép toán quan hệ hoặc logic cụ thể và tạo ra kết quả cuối cùng

Ngôn ngữ C có nhiều toán tử mạnh mẽ. Nhiều toán tử trong C là toán tử nhị phân, có nghĩa là chúng có hai toán hạng. Ví dụ, trong biểu thức a / b, / là một toán tử nhị phân nhận hai toán hạng (a, b). Ngoài ra, còn có một số toán tử một ngôi (unary operators) chỉ yêu cầu một toán hạng (ví dụ: ~, ++), và chỉ có một toán tử 3 ngôi duy nhất ?

## PHẦN 4.1 TOÁN TỬ QUAN HỆ

Các toán tử quan hệ trong ngôn ngữ C kiểm tra xem một mối quan hệ cụ thể giữa hai toán hạng có đúng hay không. Kết quả được đánh giá là 1 (đúng) hoặc 0 (sai). Kết quả này thường được sử dụng để ảnh hưởng đến luồng điều khiển (thông qua câu lệnh if, while, for), nhưng cũng có thể được lưu trữ trong các biến.

### Bằng “==”

Kiểm tra xem các toán hạng được cung cấp có bằng nhau không.

|  |
| --- |
| 1 == 0; /\* đánh giá thành 0.(Sai) \*/  1 == 1; /\* đánh giá thành 1.(Đúng) \*/  int x = 5;  int y = 5;  int \*xptr = &x, yptr = &y;  xptr == yptr; /\* đánh giá thành 0, các toán hạng giữ các địa chỉ vùng nhớ khác nhau.(Sai) \*/  \*xptr == yptr; /\* đánh giá thành 1, các toán hạng trỏ tới các vị trí vùng nhớ chứa cùng một giá trị.(Đúng) \*/ |

**Chú ý:** Không nên nhầm lẫn toán tử này với toán tử gán “=” !!!

### Không bằng (Khác) “!=”

Kiểm tra xem các toán hạng được cung cấp có bằng nhau không

|  |
| --- |
| 1 == 0; /\* đánh giá thành 0.(Sai) \*/  1 == 1; /\* đánh giá thành 1.(Đúng) \*/  int x = 5;  int y = 5;  int \*xptr = &x, yptr = &y;  xptr == yptr; /\* đánh giá thành 0, các toán hạng giữ các địa chỉ vùng nhớ khác nhau.(Sai) \*/  \*xptr == yptr; /\* đánh giá thành 1, các toán hạng trỏ tới các vị trí vùng nhớ chứa cùng một giá trị.(Đúng) \*/ |

Toán tử này trả về kết quả ngược lại với kết quả của toán tử **bằng bằng** (==).

### Not “!”

Kiểm tra xem một đối tượng có bằng 0 hay không.

Toán tử "!" cũng có thể được sử dụng trực tiếp với một biến như sau:

|  |
| --- |
| !someVal |

Điểu này có tác dụng tương tự như

|  |
| --- |
| someVal == 0 |

### So sánh lớn hơn “>”

Kiểm tra xem toán hạng bên trái có giá trị lớn hơn toán hạng bên phải hay không

|  |
| --- |
| 5 > 4 /\* trả về 1.(Đúng) \*/  4 > 5 /\* trả về 0.(Sai) \*/   1. > 4 /\* trả về 0.(Sai) \*/ |

### 4.1.5 So sánh bé hơn “<”

Kiểm tra xem toán hạng bên trái có giá trị bé hơn toán hạng bên phải hay không

|  |
| --- |
| 5 < 4 /\* trả về 0.(Sai) \*/  4 < 5 /\* trả về 1.(Đúng) \*/   1. < 4 /\* trả về 0.(Sai) \*/ |

### 4.1.6 So sánh lớn hơn hoặc bằng “>=”

Kiểm tra xem toán hạng bên trái có giá trị lớn hơn hoặc bằng toán hạng bên phải hay không

|  |
| --- |
| 5 >= 4 /\* trả về 1.(Đúng) \*/  4 >= 5 /\* trả về 0.(Sai) \*/   1. >= 4 /\* trả về 1.(Đúng) \*/ |

### 4.1.7 So sánh bé hơn hoặc bằng “<=”

Kiểm tra xem toán hạng bên trái có giá trị bé hơn hoặc bằng toán hạng bên phải hay không

|  |
| --- |
| 5 <= 4 /\* trả về 0.(Sai) \*/  4 <= 5 /\* trả về 1.(Đúng) \*/  4 <= 4 /\* trả về 1.(Đúng) \*/ |

## PHẦN 4.2 TOÁN TỬ ĐIỀU KIỆN/ TOÁN TỬ 3 NGÔI

Toán tử 3 ngôi "?" trong ngôn ngữ C được sử dụng để thực hiện một phép toán 3 ngôi (ternary operation). Nó đánh giá toán hạng đầu tiên và nếu giá trị kết quả khác 0, nó đánh giá toán hạng thứ hai. Ngược lại, nếu giá trị kết quả bằng 0, nó đánh giá toán hạng thứ ba.

|  |
| --- |
| a = b ? c : d;  /\*Tương đương với\*/  if (b)      a = c;  else      a = d; |

Đoạn mã sau đây mô tả cú pháp của toán tử 3 ngôi: **điều kiện ? kết quả nếu đúng : kết quả nếu sai**. Mỗi giá trị có thể là kết quả của một biểu thức đã được đánh giá.

|  |
| --- |
| int x = 5;      int y = 42;      printf("%i, %i\n", 1 ? x : y, 0 ? x : y); /\* Kết quả "5, 42" \*/ |

Toán tử điều kiện có thể được lồng vào nhau. Ví dụ: đoạn mã sau xác định số lớn hơn trong ba số:

|  |
| --- |
| big= a > b  ? (a > c ? a : c)              : (b > c ? b : c); |

Ví dụ sau ghi số nguyên chẵn vào một tệp và số nguyên lẻ vào tệp khác

|  |
| --- |
| #include<stdio.h>  int main()  {  *FILE* \*even, \*odds;      int n = 10;  *size\_t* k = 0;      even = fopen("even.txt", "w");      odds = fopen("odds.txt", "w");      for(k = 1; k < n + 1; k++)      {          k%2==0 ? fprintf(even, "\t%5d\n", k)                  : fprintf(odds, "\t%5d\n", k);      }      fclose(even);      fclose(odds);      return 0;  } |

Toán tử điều kiện liên kết từ phải sang trái. Hãy xem xét những điều sau đây:

|  |
| --- |
| exp1 ? exp2 : exp3 ? exp4 : exp5 |

Vì liên kết từ phải sang trái, biểu thức trên được đánh giá là:

|  |
| --- |
| exp1 ? exp2 : ( exp3 ? exp4 : exp5 ) |

## PHẦN 4.3 TOÁN TỬ BITWISE

Toán tử bitwise có thể được sử dụng để thực hiện thao tác cấp độ bit trên các biến. Dưới đây là danh sách tất cả sáu toán tử bitwise được hỗ trợ trong C:

|  |  |
| --- | --- |
| Kí hiệu | Toán tử |
| & | Toán tử AND |
| | | Toán tử OR |
| ^ | Toán tử XOR |
| ~ | Toán tử phủ định NOT |
| << | Toán tử dịch trái |
| >> | Toán tử dịch phải |

Chương trình sau đây minh họa việc sử dụng tất cả các toán tử bitwise:

|  |
| --- |
| #include <stdio.h>  int main()  {      unsigned int a = 29; /\* 29 = 0001 1101 \*/      unsigned int b = 48; /\* 48 = 0011 0000 \*/      int c = 0;      c = a & b; /\* 32 = 0001 0000 \*/      printf("%d & %d = %d\n", a, b, c );      c = a | b; /\* 61 = 0011 1101 \*/      printf("%d | %d = %d\n", a, b, c );      c = a ^ b; /\* 45 = 0010 1101 \*/      printf("%d ^ %d = %d\n", a, b, c );      c = ~a; /\* -30 = 1110 0010 \*/      printf("~%d = %d\n", a, c );      c = a << 2; /\* 116 = 0111 0100 \*/      printf("%d << 2 = %d\n", a, c );      c = a >> 2; /\* 7 = 0000 0111 \*/      printf("%d >> 2 = %d\n", a, c );      return 0;  } |

**Kết quả:**

|  |
| --- |
| **29 & 48 = 16**  **29 | 48 = 61**  **29 ^ 48 = 45**  **~29 = -30**  **29 << 2 = 116**  **29 >> 2 = 7** |

Các phép toán bitwise với kiểu dữ liệu có dấu nên được sử dụng cẩn thận vì bit dấu của kiểu dữ liệu đó có ý nghĩa riêng. Có các hạn chế đặc biệt áp dụng cho các toán tử dịch chuyển (shift operators):

* Dịch trái một bit 1 vào bit dấu của kiểu dữ liệu có dấu được coi là sai và dẫn đến hành vi không xác định. Điều này có nghĩa rằng kết quả của phép dịch như vậy không được xác định rõ ràng và có thể khác nhau tùy thuộc vào trình biên dịch và nền tảng.
* Dịch phải một giá trị âm (với bit dấu được đặt là 1) được xác định bởi từng trình biên dịch và do đó không có tính chất di động. Điều này có nghĩa rằng hành vi của phép dịch như vậy có thể khác nhau giữa các phiên bản khác nhau của ngôn ngữ C và không được đảm bảo di động trên các nền tảng hoặc trình biên dịch khác nhau.
* Nếu giá trị của toán hạng phải của toán tử dịch chuyển là số âm hoặc lớn hơn hoặc bằng độ rộng của toán hạng trái được gán, hành vi là không xác định. Điều này có nghĩa là kết quả của phép dịch như vậy không được chỉ định và có thể cho ra kết quả không mong đợi hoặc không nhất quán.

Để đảm bảo tính di động và tránh hành vi không xác định, thường khuyến nghị sử dụng kiểu dữ liệu không dấu khi thực hiện các phép toán bitwise liên quan đến dịch chuyển.

**Masking:**

Masking (hoặc còn gọi là áp dụng mặt nạ) đề cập đến quá trình trích xuất các bit mong muốn từ một biến (hoặc biến đổi các bit mong muốn trong biến) bằng cách sử dụng các phép toán logic bitwise. Toán hạng (một hằng số hoặc biến) được sử dụng để thực hiện việc áp dụng mặt nạ được gọi là mặt nạ (mask).

Trong quá trình masking, mặt nạ là một giá trị có các bit được thiết lập một cách đặc biệt để thực hiện việc trích xuất hoặc biến đổi các bit trong biến gốc. Khi thực hiện phép toán bitwise AND (&) giữa biến gốc và mặt nạ, chỉ có các bit tương ứng mà mặt nạ đặt thành 1 và các bit còn lại trong biến gốc sẽ bị xóa đi (thiết lập thành 0). Điều này cho phép trích xuất hoặc biến đổi các bit cần thiết trong biến gốc và bỏ qua các bit không mong muốn.

Mặt nạ được sử dụng theo nhiều cách khác nhau

* Để quyết định mẫu bit của một biến số nguyên.
* Để sao chép một phần của mẫu bit đã cho sang một biến mới, trong khi phần còn lại của biến mới được điền bằng 0 (sử dụng bitwise AND)
* Để sao chép một phần của mẫu bit đã cho sang một biến mới, trong khi phần còn lại của biến mới được lấp đầy bằng 1 (sử dụng OR theo bit).
* Để sao chép một phần của mẫu bit đã cho sang một biến mới, trong khi phần còn lại của mẫu bit ban đầu được đảo ngược trong biến mới (sử dụng OR loại trừ bit).

Hàm sau sử dụng mặt nạ để hiển thị mẫu bit của một biến:

|  |
| --- |
| #include <limits.h>  void bit\_pattern(int *u*)  {      int i, x, word;      unsigned mask = 1;      word = CHAR\_BIT \* sizeof(int);      mask = mask << (word - 1); /\* shift 1 to the leftmost position \*/      for(i = 1; i <= word; i++)      {          x = (*u* & mask) ? 1 : 0; /\* identify the bit \*/          printf("%d", x); /\* print bit value \*/          mask >>= 1; /\* shift mask to the right by 1 bit \*/      }  } |

## PHẦN 4.4 HOẠT ĐỘNG NGẮN MẠCH CỦA TOÁN TỬ LOGIC

## PHẦN 4.5 TOÁN TỬ DẤU PHẨY

Toán tử dấu phẩy (,) cho phép bạn đánh giá nhiều biểu thức ở bất kỳ nơi nào cho phép một biểu thức duy nhất. Toán tử dấu phẩy đánh giá toán hạng bên trái, sau đó là toán hạng bên phải, rồi trả về kết quả của toán hạng bên phải.

|  |
| --- |
| int x = 42, y = 42;  printf("%i\n", (x \*= 2, y)); /\* Outputs "42". \*/ |

Toán tử dấu phẩy giới thiệu một điểm thứ tự giữa các toán hạng của nó.

**Lưu ý** rằng dấu phẩy được sử dụng trong các hàm gọi các đối số riêng biệt đó **KHÔNG** phải là toán tử dấu phẩy, thay vào đó, nó được gọi là dấu phân cách khác với toán tử dấu phẩy. Do đó, nó không có các thuộc tính của toán tử dấu phẩy.

Lệnh gọi printf() ở trên chứa cả toán tử dấu phẩy và dấu phân cách.

|  |
| --- |
| printf("%i\n", (x \*= 2, y)); /\* Outputs "42". \*/  /\*           ^        ^ Đây là toán tử dấu phẩy \*/  /\*           Đây là dấu phân cách \*/ |

Toán tử dấu phẩy thường được sử dụng trong phần khởi tạo cũng như trong phần cập nhật của vòng lặp for. Ví dụ:

|  |
| --- |
| for(k = 1; k < 10; printf("*\%*d\\n", k), k += 2); /\*outputs the odd numbers below 9/\*  /\* outputs sum to first 9 natural numbers \*/  for(sumk = 1, k = 1; k < 10; k++, sumk += k)      printf("*\%*5d*\%*5d\\n", k, sumk); |

## PHẦN 4.6 TOÁN TỬ SỐ HỌC

### 4.6.1 Số học cơ bản

Trả về một giá trị là kết quả của việc áp dụng toán hạng bên trái cho toán hạng bên phải, sử dụng phép toán liên quan. Các quy tắc giao hoán toán học thông thường được áp dụng (nghĩa là phép cộng và phép nhân có tính chất giao hoán, phép trừ, phép chia và mô đun thì không).

### 4.6.2 Toán tử cộng

Toán tử cộng (+) được sử dụng để cộng hai toán hạng lại với nhau. Ví dụ:

|  |
| --- |
| #include<stdio.h>  int main(){      int a = 5;      int b = 7;      int c = a + b;      printf("%d + %d = %d",a,b,c);      return 0;  } |

**Kết quả**

|  |
| --- |
| **5 + 7 = 12** |

### 4.6.3 Toán tử trừ

Toán tử trừ được sử dụng để trừ toán hạng thứ 2 cho toán hạng thứ nhất. Ví dụ

|  |
| --- |
| #include<stdio.h>  int main(){      int a = 5;      int b = 7;      int c = a - b;      printf("%d - %d = %d",a,b,c);      return 0;  } |

**Kết quả**

|  |
| --- |
| **5 - 7 = -2** |

### 4.6.4 Toán tử nhân

Toán tử nhân (\*) được sử dụng để nhân cả hai toán hạng. Ví dụ:

|  |
| --- |
| #include<stdio.h>  int main(){      int a = 5;      int b = 7;      int c = a \* b;      printf("%d \* %d = %d",a,b,c);      return 0;  } |

**Kết quả**

|  |
| --- |
| **5 \* 7 = 35** |

**Đừng nhầm lẫn với con trỏ (\*)**

### 4.6.5 Toán tử chia

Toán tử chia (/) chia toán hạng thứ nhất cho toán hạng thứ hai. Nếu cả hai toán hạng của phép chia là số nguyên, nó sẽ trả về một giá trị nguyên và loại bỏ phần còn lại (sử dụng toán tử modulo % để tính toán và lấy phần còn lại).

Nếu một trong các toán hạng là một giá trị dấu chấm động, thì kết quả là một phép tính gần đúng của phân số.

**Ví dụ:**

|  |
| --- |
| #include<stdio.h>  int main(){      int a = 19 / 2 ;      int b = 18 / 2 ;      int c = 255 / 2;      int d = 44 / 4 ;      double e = 19 / 2.0 ;      double f = 18.0 / 2 ;      double g = 255 / 2.0;      double h = 45.0 / 4 ;      printf("19 / 2 = %d\n", a);      printf("18 / 2 = %d\n", b);      printf("255 / 2 = %d\n", c);      printf("44 / 4 = %d\n", d);      printf("19 / 2.0 = %g\n", e);      printf("18.0 / 2 = %g\n", f);      printf("255 / 2.0 = %g\n", g);      printf("45.0 / 4 = %g\n", h);  } |

**Kết quả**

|  |
| --- |
| **19 / 2 = 9**  **18 / 2 = 9**  **255 / 2 = 127**  **44 / 4 = 11**  **19 / 2.0 = 9.5**  **18.0 / 2 = 9**  **255 / 2.0 = 127.5**  **45.0 / 4 = 11.25** |

**Chú ý: Ép kiểu**

|  |
| --- |
| #include<stdio.h>  int main(){      int a = 9;      int b = 2;      float c;      //c = (float) a / b;      //c = 9.0 / 2;      c = 9 / 2.0;      printf("%d / %d = %f",a,b,c);  } |

**Kết quả**

|  |
| --- |
| **9 / 2 = 4.500000** |

### 4.6.6 Toán tử modulo (Phép chia lấy dư)

Toán tử modulo (%) chỉ nhận toán hạng số nguyên và được sử dụng để tính phần còn lại sau khi toán hạng thứ nhất được chia cho toán hạng thứ hai. Toán tử mô đun (còn được gọi là toán tử chia lấy phần dư) là một toán tử trả về phần còn lại sau khi thực hiện phép chia số nguyên.

Ví dụ:

|  |
| --- |
| #include <stdio.h>  int main (void) {      int a = 25 % 2; /\* a holds value 1 \*/      int b = 24 % 2; /\* b holds value 0 \*/      int c = 155 % 5; /\* c holds value 0 \*/      int d = 49 % 25; /\* d holds value 24 \*/      printf("25 *%* 2 = %d\n", a); /\* Will output "25 % 2 = 1" \*/      printf("24 *%* 2 = %d\n", b); /\* Will output "24 % 2 = 0" \*/      printf("155 *%* 5 = %d\n", c); /\* Will output "155 % 5 = 0" \*/      printf("49 *%* 25 = %d\n", d); /\* Will output "49 % 25 = 24" \*/      return 0;  } |

### 4.6.7 Toán tử tăng giảm

Các toán tử tăng (a++) và giảm (a--) khác nhau ở chỗ chúng thay đổi giá trị của biến mà bạn áp dụng chúng mà không có toán tử gán. Bạn có thể sử dụng toán tử **tăng** và **giảm** trước hoặc sau biến. Vị trí của toán tử thay đổi thời gian tăng/giảm giá trị thành trước hoặc sau khi gán nó cho biến.

Ví dụ:

|  |
| --- |
| #include <stdio.h>  int main(void)  {      int a = 1;      int b = 4;      int c = 1;      int d = 4;      a++;      printf("a = %d\n",a); /\* Sẽ in ra "a = 2" \*/      b--;      printf("b = %d\n",b); /\* Sẽ in ra "b = 3" \*/      if (++c > 1) { /\* c được tăng thêm 1 trước khi được so sánh trong điều kiện \*/          printf("This will print\n"); /\* Dòng này sẽ được in ra \*/      } else {          printf("This will never print\n"); /\* Dòng này sẽ không được in ra \*/      }      if (d-- < 4) { /\* d được giảm sau khi được so sánh \*/          printf("This will never print\n"); /\* Dòng này sẽ không được in ra \*/      } else {          printf("This will print\n"); /\* Dòng này sẽ được in ra \*/      }  } |

Kết quả

|  |
| --- |
| a = 2  b = 3  This will print  This will print |

Như ví dụ cho c và d cho thấy, cả hai toán tử đều có hai dạng, là ký hiệu tiền tố và ký hiệu hậu tố. Cả hai đều có tác dụng giống nhau trong việc tăng (++) hoặc giảm (--) biến, nhưng khác nhau về giá trị mà chúng trả về: các phép toán tiền tố thực hiện thao tác trước rồi trả về giá trị, trong khi các phép toán hậu tố trước tiên xác định giá trị được trả lại, và sau đó thực hiện thao tác.

Do hành vi có khả năng phản trực giác này, việc sử dụng các toán tử tăng/giảm bên trong các biểu thức đang gây tranh cãi.

## PHẦN 4.7 TOÁN TỬ TRUY CẬP

Các toán tử truy cập thành viên (dấu chấm . và mũi tên ->) được sử dụng để truy cập một thành viên của cấu trúc(struct).

### 4.7.1 Thuộc tính của đối tượng

Được đánh giá thành giá trị, chỉ định đối tượng là thành viên của đối tượng được truy cập.

|  |
| --- |
| struct *MyStruct*  {      int x;      int y;  };  struct *MyStruct* myObject;  myObject.x = 42;  myObject.y = 123;  printf(".x = %i, .y = %i\n", myObject.x, myObject.y); /\* Outputs ".x = 42, .y = 123". \*/ |

### 4.7.2 Thành viên của đối tượng được trỏ tới

Toán tử mũi tên (->) được sử dụng để tiện lợi hóa quá trình giải tham chiếu và truy cập thành viên của một cấu trúc. Thay vì viết (\*x).y để truy cập thành viên y của đối tượng được giải tham chiếu bởi con trỏ x, chúng ta có thể sử dụng x->y. Điều này giúp làm rõ hơn cách thao tác và làm cho mã nguồn dễ đọc hơn, đặc biệt là khi có các con trỏ cấu trúc lồng nhau.

|  |
| --- |
| #include <stdio.h>  int main(){  struct *MyStruct*  {      int x;      int y;  };  struct *MyStruct* myObject;  struct *MyStruct* \*p = &myObject;  p->x = 42;  p->y = 123;  printf(".x = %i, .y = %i\n", p->x, p->y); /\* Kết quả ".x = 42, .y = 123". \*/  printf(".x = %i, .y = %i\n", myObject.x, myObject.y); /\* Kết quả cũng sẽ là ".x = 42, .y = 123". \*/  } |

### 4.7.3 Lấy địa chỉ

Toán tử & (unary address of operator) trong ngôn ngữ C được sử dụng để lấy địa chỉ của một biến. Khi áp dụng toán tử & cho một biến, nó trả về một con trỏ chứa địa chỉ của biến đó.

**Ví dụ:**

|  |
| --- |
| #include <stdio.h>  int main(){  int x = 3;  int \*p = &x;  printf("%p = %p\n", &x, (p)); /\* Xuất ra "A = A",Với A là một giá trị do triển khai cụ thể quy định ở đây là giá trị địa chỉ ngẫu nhiên.\*/  } |

**Kết quả**

|  |
| --- |
| 0061FF18 = 0061FF18 |

### 4.7.4 Dereference

The unary \* operator dereferences a pointer. It evaluates into the lvalue resulting from dereferencing the pointer that results from evaluating the given expression.

|  |
| --- |
| int x = 42;  int \*p = &x;  printf("x = %d, \*p = %d\n", x, \**p*); /\* Xuất ra "x = 42, \*p = 42". \*/  \*p = 123;  printf("x = %d, \*p = %d\n", x, \**p*); /\* Xuất ra "x = 123, \*p = 123". \*/ |

### 4.7.5 Chỉ số mảng

Trong C, cú pháp indexing (chỉ số mảng) thực tế là một cách viết tắt cho việc thực hiện phép cộng con trỏ và sau đó giải tham chiếu. Biểu thức dạng **a[i]** tương đương với **\*(a + i)**, trong đó a là một con trỏ và i là chỉ số hoặc độ lệch.

Trong C, mảng sẽ tự động chuyển thành con trỏ tới phần tử đầu tiên của nó trong hầu hết các ngữ cảnh. Khi bạn sử dụng cú pháp chỉ số **a[i],** trình biên dịch sẽ coi a như một con trỏ tới phần tử đầu tiên của mảng và thực hiện phép cộng con trỏ với độ lệch **i**. Sau đó, con trỏ kết quả được giải tham chiếu sử dụng toán tử **\*** để truy cập vào giá trị tại vị trí đó trong bộ nhớ.

|  |
| --- |
| int arr[] = { 1, 2, 3, 4, 5 };  printf("arr[2] = %i\n", arr[2]); /\* Xuất ra "arr[2] = 3". \*/ |

### 4.7.6 Khả năng hoán đổi của chỉ số mảng

Thêm một con trỏ vào một số nguyên là một phép toán giao hoán (nghĩa là thứ tự của các toán hạng không làm thay đổi kết quả) vì vậy con trỏ + số nguyên == số nguyên + con trỏ.

Hệ quả của điều này là arr[3] và 3[arr] là tương đương.

|  |
| --- |
| printf("3[arr] = %i\n", 3[arr]); /\* Xuất ra "3[arr] = 4". \*/ |

## PHẦN 4.8 TOÁN TỬ SIZEOF

### 4.8.1 With a type as operand

Đánh giá kết quả là kích thước trong byte, của kiểu size\_t, của đối tượng thuộc kiểu đã cho. Yêu cầu có dấu ngoặc đơn xung quanh kiểu dữ liệu.

|  |
| --- |
| printf("%zu\n", sizeof(int)); /\* Hợp lệ, in ra kích thước của một đối tượng int,                                 \*có thể thay đổi theo nền tảng. \*/  printf("%zu\n", sizeof *int*); /\* Không hợp lệ, các kiểu dữ liệu truyền vào cần được                                \*bao bọc bởi dấu ngoặc đơn! \*/ |

### 4.8.2 With an expression as operand

Đánh giá kết quả là kích thước trong byte, kiểu size\_t, của các đối tượng có cùng kiểu dữ liệu với biểu thức được cung cấp. Biểu thức chính nó không được đánh giá. Dấu ngoặc đơn không bắt buộc; tuy nhiên, vì biểu thức được cung cấp phải là toán tử một ngôi, nên luôn tốt nhất là sử dụng chúng.

|  |
| --- |
| char ch = 'a';  printf("%zu\n", sizeof(ch)); /\* Hợp lệ, sẽ in ra kích thước của một đối tượng char,                                \* luôn là 1 cho mọi nền tảng. \*/  printf("%zu\n", sizeof *ch*); /\* Hợp lệ, sẽ in ra kích thước của một đối tượng char,                               \*luôn là 1 cho mọi nền tảng. \*/ |

## PHẦN 4.9 TOÁN TỬ ÉP KIỂU

Thực hiện một chuyển đổi rõ ràng sang kiểu dữ liệu đã chỉ định từ giá trị thu được sau khi đánh giá biểu thức cung cấp.

|  |
| --- |
| int x = 3;  int y = 4;  printf("%f\n", (double)x / y); /\* In ra "0.750000". \*/ |

Ở đây, giá trị của x được chuyển đổi thành kiểu double, phép chia cũng chuyển đổi giá trị của y thành kiểu double và kết quả của phép chia, một giá trị double, được truyền vào hàm printf để in ra màn hình.

## PHẦN 4.10 TOÁN TỬ GỌI HÀM

Toán tử gọi hàm trong ngôn ngữ lập trình yêu cầu toán hạng đầu tiên phải là một con trỏ tới hàm (function pointer) hoặc một biểu thức chỉ định hàm (function designator). Đối tượng này xác định hàm mà ta muốn gọi. Các toán hạng còn lại, nếu có, được gọi là đối số của cuộc gọi hàm. Toán tử này đánh giá thành giá trị trả về từ việc gọi hàm thích hợp với các đối số tương ứng.

|  |
| --- |
| int myFunction(int *x*, int *y*)  {      return *x* \* 2 + *y*;  }  int (\*fn)(int, int) = &myFunction;  int x = 42;  int y = 123;  printf("(\*fn)(%i, %i) = %i\n", x, y, (fn)(x, y)); /\* In ra "fn(42, 123) = 207". \*/  printf("fn(%i, %i) = %i\n", x, y, fn(*x*, y)); /\* Một dạng khác: bạn không cần phải giải tham chiếu một cách rõ ràng \*/ |

## PHẦN 4.11 TOÁN TỬ TĂNG/ TOÁN TỬ GIẢM

Các toán tử tăng và giảm tồn tại ở dạng tiền tố và hậu tố.

|  |
| --- |
| int a = 1;  int b = 1;  int tmp = 0;  tmp = ++a; /\* tăng a lên một đơn vị và trả về giá trị mới; a == 2, tmp == 2 \*/  tmp = a++; /\* tăng a lên một đơn vị nhưng trả về giá trị cũ; a == 3, tmp == 2 \*/  tmp = --b; /\* giảm b đi một đơn vị và trả về giá trị mới; b == 0, tmp == 0 \*/  tmp = b--; /\* giảm b đi một đơn vị nhưng trả về giá trị cũ; b == -1, tmp == 0 \*/ |

Hiểu đơn giản

Toán tử ++: Tăng giá trị lên 1 đơn vị.

**• x++:** thực hiện lệnh trước rồi mới tăng x lên 1 đơn vị.

**• ++x:** tăng x lên 1 đơn vị rồi mới thực hiện lệnh.

Toán tử --: Giảm giá trị đi 1 đơn vị.

**• x--:** thực hiện lệnh trước rồi mới giảm x đi 1 đơn vị.

**• --x:** giảm x đi 1 đơn vị rồi mới thực hiện lệnh.

Lưu ý rằng các phép toán số học không tạo ra điểm chuỗi (sequence points), do đó, một số biểu thức sử dụng toán tử tăng (++) hoặc giảm (--) có thể dẫn đến hành vi không xác định.

## PHẦN 4.12 TOÁN TỬ GÁN

Gán giá trị của toán hạng bên phải cho vị trí lưu trữ được đặt tên bởi toán hạng bên trái và trả về giá trị.

|  |
| --- |
| int x = 5; /\* Biến x chứa giá trị 5. Trả về 5. /  char y = 'c'; /\* Biến y chứa giá trị 99. Trả về 99                 \*(vì ký tự 'c' được đại diện bởi 99 trong bảng ASCII). \*/  float z = 1.5; /\* Biến z chứa giá trị 1.5. Trả về 1.5. \*/  char const\* s = "foo"; /\* Biến s chứa địa chỉ của ký tự đầu tiên trong chuỗi 'foo'. \*/ |

Một số phép tính số học có toán tử gán phức hợp.

|  |
| --- |
| a += b /\* tương đương với: a = a + b \*/  a -= b /\* tương đương với: a = a - b \*/  a = b /\* tương đương với: a = a \* b \*/  a /= b /\* tương đương với: a = a / b \*/  a %= b /\*tương đương với: a = a % b \*/  a &= b /\* tương đương với: a = a & b \*/  a |= b /\* tương đương với: a = a | b \*/  a ^= b /\* tương đương với: a = a ^ b \*/  a <<= b /\* tương đương với: a = a << b \*/  a >>= b /\*tương đương với: a = a >> b \*/ |

Một tính năng quan trọng của các toán tử gán kết hợp này là biểu thức ở phía bên trái (a) chỉ được đánh giá một lần. Ví dụ, nếu p là một con trỏ:

|  |
| --- |
| \*p += 5; |

Trong trường hợp này, giá trị của biểu thức \*p sẽ chỉ được tính toán một lần và sau đó, giá trị mới được gán vào vị trí mà con trỏ p trỏ tới. Điều này giúp tránh các lỗi không xác định hoặc không mong muốn khi giá trị của biểu thức thay đổi trong quá trình đánh giá.

|  |
| --- |
| \*p = \*p + 5 |

Cũng cần lưu ý rằng kết quả của một phép gán chẳng hạn như a = b là cái được gọi là giá trị. Do đó, phép gán thực sự có một giá trị mà sau đó có thể được gán cho một biến khác. Điều này cho phép xâu chuỗi các phép gán để đặt nhiều biến trong một câu lệnh.

Giá trị này có thể được sử dụng trong các biểu thức kiểm soát của câu lệnh if (hoặc vòng lặp hoặc câu lệnh chuyển đổi) để bảo vệ một số mã trên kết quả của một biểu thức hoặc lời gọi hàm khác. Ví dụ:

|  |
| --- |
| char \*buffer;  if ((buffer = malloc(1024)) != NULL)  {   /\* thực hiện một số công việc sử dụng buffer \*/      free(buffer);  }  else  {   /\* báo cáo lỗi khi không cấp phát được bộ nhớ \*/  } |

Do đó, cần phải cẩn thận để tránh mắc lỗi đánh máy phổ biến có thể dẫn đến các lỗi bí ẩn.

|  |
| --- |
| int a = 2;  /\* ... \*/  if (a = 1)  /\* Xóa tất cả các tập tin trên ổ cứng của tôi \*/ |

Điều này sẽ dẫn đến kết quả tai hại, vì a = 1 sẽ luôn đánh giá bằng 1 và do đó, biểu thức kiểm soát của câu lệnh if sẽ luôn đúng (đọc thêm về cạm bẫy phổ biến này tại đây). Tác giả gần như chắc chắn có ý định sử dụng toán tử đẳng thức (==) như hình dưới đây:

|  |
| --- |
| int a = 2;  /\* ... \*/  if (a == 1)  /\* Xóa tất cả các tập tin trên ổ cứng của tôi \*/ |

**Tính kết hợp của toán tử**

|  |
| --- |
| int a, b = 1, c = 2;  a = b = c; |

Điều này gán c cho b, trả về b, được gán cho a. Điều này xảy ra bởi vì tất cả các toán tử gán có tính kết hợp phải, điều đó có nghĩa là phép toán ngoài cùng bên phải trong biểu thức được ước tính trước và tiến hành từ phải sang trái.

## PHẦN 4.13 TOÁN TỬ LOGIC

### 4.13.1 Logic AND

Toán tử logic AND (&&) trong C thực hiện phép AND logic giữa hai toán hạng và trả về giá trị 1 nếu cả hai toán hạng đều khác 0. (Có nghĩa là đúng khi cả 2 đều đúng) Toán tử AND logic có kiểu int.

|  |
| --- |
| 0 && 0 /\* Trả về 0. \*/  0 && 1 /\* Trả về 0. \*/  2 && 0 /\* Trả về 0. \*/  2 && 3 /\* Trả về 1. \*/ |

### 4.13.2 Logic OR

Toán tử logic OR (||) trong C thực hiện phép OR logic giữa hai toán hạng và trả về giá trị 1 nếu bất kỳ một trong hai toán hạng là khác không. Toán tử OR logic có kiểu int.

|  |
| --- |
| 0 || 0 /\* Trả về 0. \*/  0 || 1 /\* Trả về 1. \*/  2 || 0 /\* Trả về 1. \*/  2 || 3 /\* Trả về 1. \*/ |

### 4.13.3 Logic NOT

Toán tử logic NOT (!) trong C thực hiện phép phủ định logic của một biểu thức và trả về giá trị 1 nếu biểu thức là sai (bằng 0), và trả về giá trị 0 nếu biểu thức là đúng (khác 0). Toán tử NOT logic có kiểu int.

Các tính chất quan trọng chung cho cả phép toán logic AND (&&) và phép toán logic OR (||) là như sau, và rất quan trọng để hiểu cách chúng hoạt động trong việc đánh giá ngắn gọn (short-circuit evaluation):

Toán hạng bên trái (LHS) được đánh giá hoàn toàn trước khi toán hạng bên phải (RHS) được đánh giá:

* Trong biểu thức A && B, A được đánh giá trước, và chỉ khi A là đúng (khác không) thì B mới được đánh giá. Nếu A là sai (bằng không), B sẽ không được đánh giá vì kết quả tổng thể của phép toán AND đã xác định là sai.
* Trong biểu thức A || B, A được đánh giá trước, và chỉ khi A là sai (bằng không) thì B mới được đánh giá. Nếu A là đúng (khác không), B sẽ không được đánh giá vì kết quả tổng thể của phép toán OR đã xác định là đúng.

Có một điểm chuỗi (sequence point) giữa việc đánh giá toán hạng bên trái và toán hạng bên phải:

* Điểm chuỗi là điểm trong việc thực thi chương trình nơi tất cả các hiệu ứng phụ của các đánh giá trước đó đã được thực hiện. Điểm chuỗi đảm bảo rằng các đánh giá xảy ra theo một thứ tự được xác định, tránh các vấn đề liên quan đến hành vi không xác định (undefined behavior).

Toán hạng bên phải không được đánh giá nếu kết quả của toán hạng bên trái xác định kết quả tổng thể:

* Đánh giá ngắn gọn (short-circuit evaluation) tránh việc thực hiện các tính toán không cần thiết. Ví dụ, trong A && B, nếu A là sai, kết quả của biểu thức tổng thể đã biết là sai bất kể giá trị của B. Do đó, không cần phải đánh giá B.
* Tương tự, trong A || B, nếu A là đúng, kết quả của biểu thức tổng thể đã biết là đúng bất kể giá trị của B. Do đó, không cần phải đánh giá B.

Các tính chất này là rất quan trọng để viết mã hiệu quả và an toàn, đặc biệt khi sử dụng các biểu thức có hiệu ứng phụ tiềm ẩn hoặc các đánh giá phức tạp. Nó cho phép các lập trình viên tối ưu hóa mã của họ và tránh công việc không cần thiết khi đánh giá biểu thức logic.

## PHẦN 4.14 TOÁN HỌC TRÊN CON TRỎ

### 4.14.1 Cộng con trỏ

### 4.14.2 Trừ con trỏ

## PHẦN 4.15 \_Alignof

# BOOLEAN

## PHẦN 5.1 SỬ DỤNG STDBOOL.H

Tiêu sử dụng tệp tiêu đề hệ thống **stdbool.h** cho phép bạn sử dụng **bool** như một kiểu dữ liệu Boolean. **true** tương đương với **1** và **false** tương đương với **0**.

|  |
| --- |
| #include <stdio.h>  #include <stdbool.h>  int main(void) {      bool x = true; /\* tương đương với bool x = 1; \*/      bool y = false; /\* tương đương với bool y = 0; \*/      if (x) /\* Tương đương với if (x != 0) hoặc if (x != false) \*/      {          puts("This will print!"); /\* In ra màn hình chuỗi "This will print!" \*/      }      if (!y) /\* Tương đương với if (y == 0) hoặc if (y == false) \*/      {          puts("This will also print!"); /\* In ra màn hình chuỗi "This will also print!" \*/      }      return 0;  } |

Trong C, từ khóa "bool" chỉ là một tên gọi khác cho kiểu dữ liệu "\_Bool". Kiểu dữ liệu "\_Bool" là một kiểu dữ liệu tích hợp được giới thiệu trong tiêu chuẩn C99 để biểu diễn giá trị boolean, chỉ có hai giá trị có thể xảy ra: đúng (1) và sai (0).

Kiểu dữ liệu "\_Bool" có các quy tắc đặc biệt khi chuyển đổi các số hoặc con trỏ sang nó.

## PHẦN 5.2 SỬ DỤNG #DEFINE

Trong C, mọi phiên bản, các toán tử so sánh thực sự sẽ xem xét bất kỳ giá trị số nguyên nào khác 0 là đúng (true) và giá trị số nguyên 0 là sai (false). Điều này áp dụng cho các biểu thức điều kiện như trong câu lệnh if, while, for và các phép so sánh khác.

Nếu bạn không có sẵn \_Bool hoặc bool như trong C99, bạn có thể mô phỏng một kiểu dữ liệu Boolean trong C bằng cách sử dụng các macro được định nghĩa bằng #define, và bạn có thể vẫn tìm thấy những điều tương tự trong mã nguồn cũ.

|  |
| --- |
| #include <stdio.h>  #define bool int  #define true 1  #define false 0  int main(void) {      bool x = true; /\* Tương đương với int x = 1; \*/      bool y = false; /\* Tương đương với int y = 0; \*/      if (x) /\* Tương đương với if (x != 0) hoặc if (x != false) \*/      {          puts("This will print!"); /\* In ra màn hình chuỗi "This will print!" \*/      }      if (!y) /\* Tương đương với if (y == 0) hoặc if (y == false) \*/      {          puts("This will also print!"); /\* In ra màn hình chuỗi "This will also print!" \*/      }      return 0;  } |

Trong mã nguồn mới, nên tránh định nghĩa những macro như bool, true, và false, vì chúng có thể gây xung đột với việc sử dụng hiện đại của thư viện <stdbool.h>.

## PHẦN 5.3 SỬ DỤNG KIỂU \_BOOL NỘI TẠI(TÍCH HỢP)

Kiểu dữ liệu \_Bool được thêm vào trong tiêu chuẩn C99 và là một kiểu dữ liệu nguyên thuỷ (native) trong ngôn ngữ C. Nó có khả năng lưu trữ các giá trị **0** (để biểu diễn **false**) và **1** (để biểu diễn **true**).

|  |
| --- |
| #include <stdio.h>  int main(void) {      \_Bool x = 1; /\* \_Bool có thể lưu trữ giá trị 0 hoặc 1 \*/      \_Bool y = 0;      if (x) /\* Tương đương với if (x == 1) \*/      {          puts("This will print!"); /\* In ra màn hình chuỗi "This will print!" \*/      }      if (!y) /\* Tương đương với if (y == 0) \*/      {          puts("This will also print!"); /\* In ra màn hình chuỗi "This will also print!" \*/      }      return 0;  } |

Trong ngôn ngữ C, \_Bool là một kiểu số nguyên, nhưng nó có các quy tắc đặc biệt cho việc chuyển đổi từ các kiểu dữ liệu khác khi được sử dụng trong biểu thức. Kết quả của các biểu thức \_Bool tương tự như việc sử dụng các kiểu số nguyên khác trong các biểu thức if. Trong ví dụ sau:

|  |
| --- |
| \_Bool z = X; |

Nếu X có kiểu số học (là bất kỳ loại số nào), thì z sẽ trở thành 0 nếu X bằng 0. Ngược lại, z sẽ trở thành 1.

Nếu X có kiểu con trỏ (kiểu dữ liệu liên quan đến con trỏ), thì z sẽ trở thành 0 nếu X là một con trỏ null (con trỏ không trỏ tới địa chỉ nào). Ngược lại, z sẽ trở thành 1 nếu X là một con trỏ hợp lệ (con trỏ trỏ tới địa chỉ nào đó).

Để sử dụng cách viết đẹp hơn với từ khóa bool, false, và true, bạn nên bao gồm tiêu đề <stdbool.h> trong mã C của mình.

## PHẦN 5.4 SỐ NGUYÊN VÀ CON TRỎ TRONG BIỂU THỨC BOOLEAN

Trong ngôn ngữ C, tất cả các số nguyên hoặc con trỏ đều có thể được sử dụng trong các biểu thức mà được hiểu là "giá trị đúng" (truth value) trong ngữ cảnh Logic.

|  |
| --- |
| int main(int *argc*, char\* *argv*[]) {      if (*argc* % 4) {          puts("arguments number is not divisible by 4");      } else {          puts("argument number is divisible by 4");      }  ... |

Biểu thức argc % 4 sẽ đánh giá thành một trong các giá trị 0, 1, 2 hoặc 3, tùy thuộc vào phần dư khi argc chia cho 4

* Nếu kết quả của argc % 4 là 0, nó sẽ được coi là "sai" (false) trong ngữ cảnh Logic (Boolean), và chương trình sẽ thực thi mã trong khối else, in ra thông báo "argument number is divisible by 4" (số lượng đối số chia hết cho 4).
* Nếu kết quả của argc % 4 là 1, 2 hoặc 3, nó sẽ được coi là "đúng" (true) trong ngữ cảnh Logic (Boolean), và chương trình sẽ thực thi mã trong khối if, in ra thông báo "arguments number is not divisible by 4" (số lượng đối số không chia hết cho 4).

|  |
| --- |
| double\* A = malloc(n \* sizeof(\*A));  if ( !A ) {      perror("allocation problems");      exit(EXIT\_FAILURE);  } |

Trong đoạn mã trên, ta cấp phát bộ nhớ động cho con trỏ A bằng hàm malloc và sau đó kiểm tra xem A có là con trỏ null hay không. Nếu A là con trỏ null, đoạn mã in ra thông báo lỗi và chương trình thoát. Điều này giúp phát hiện và xử lý vấn đề về cấp phát bộ nhớ một cách chính xác.

|  |
| --- |
| char const\* s = ....; /\* some pointer that we receive \*/  if (s != NULL && s[0] != '\0' && isalpha(s[0])) {   printf("this starts well, %c is alphabetic\n", s[0]);  } |

Để kiểm tra điều này, bạn phải quét mã phức tạp trong biểu thức và chắc chắn về tùy chọn toán tử.

|  |
| --- |
| char const\* s = ....; /\* some pointer that we receive \*/  if (s && s[0] && isalpha(s[0])) {   printf("this starts well, %c is alphabetic\n", s[0]);  } |

tương đối dễ nắm bắt: nếu con trỏ hợp lệ, chúng tôi kiểm tra xem ký tự đầu tiên có khác 0 hay không và sau đó kiểm tra xem đó có phải là một chữ cái hay không.

## PHẦN 5.5 XÁC ĐỊNH KIỂU BOOL BẰNG CÁCH SỬ DỤNG TYPEDEF

Thay vì sử dụng các **#define** macros, ta có thể sử dụng **enum** để định nghĩa các hằng số có tên để biểu thị giá trị true và false.

|  |
| --- |
| #if \_\_STDC\_VERSION\_\_ < 199900L  typedef enum { false, true } bool;  /\* Modern C code might expect these to be macros. \*/  # ifndef bool  # define bool bool  # endif  # ifndef true  # define true true  # endif  # ifndef false  # define false false  # endif  #else  # include <stdbool.h>  #endif  /\* Somewhere later in the code ... \*/  bool b = true; |

Điều này cho phép trình biên dịch cho các phiên bản C lịch sử hoạt động, nhưng vẫn tương thích về phía trước nếu mã được biên dịch bằng trình biên dịch C hiện đại.

Để biết thêm thông tin về typedef, hãy xem Typedef, để biết thêm về enum, hãy xem Enumerations

# CHUỖI

Trong ngôn ngữ C, chuỗi (string) không phải là một kiểu dữ liệu tự nhiên như số nguyên hay số thực. Thay vào đó, một chuỗi C được biểu diễn dưới dạng một mảng một chiều các ký tự (character array), kết thúc bằng ký tự null (null-character) là '\0'.

Điều này có nghĩa là một chuỗi C có nội dung là "abc" sẽ có bốn ký tự 'a', 'b', 'c' và '\0'.

Xem ví dụ giới thiệu cơ bản về chuỗi.

## PHẦN 6.1 PHÂN TÁCH TỪ: STRTOK(), STRTOK\_R() VÀ STRTOK\_S()

Hàm **strtok()** trong ngôn ngữ C được sử dụng để chia một chuỗi thành các chuỗi nhỏ hơn hoặc các phần tử con gọi là "**tokens**" bằng cách sử dụng một tập hợp các ký tự phân tách.

**Ví dụ:**

|  |
| --- |
| #include <stdio.h>  #include <string.h>  int main(void)  {      int toknum = 0;      char src[] = "Hello,, world!";      const char delimiters[] = ", !";      char \*token = strtok(src, delimiters);      while (token != NULL)      {          printf("%d: [%s]\n", ++toknum, token);          token = strtok(NULL, delimiters);      }      /\* source is now "Hello\0, world\0\0" \*/  } |

**Kết quả:**

|  |
| --- |
| 1: [Hello]  2: [world] |

Chuỗi ký tự phân tách có thể chứa một hoặc nhiều ký tự phân tách, và bạn có thể sử dụng các chuỗi phân tách khác nhau trong mỗi lần gọi strtok().

Để tiếp tục phân tách chuỗi nguồn mà không truyền lại chuỗi nguồn, bạn nên truyền NULL làm đối số đầu tiên. Nếu truyền lại chuỗi nguồn, lần gọi đầu tiên của **strtok()** sẽ trả về lại token đầu tiên.

Hàm **strtok()** không cấp phát bộ nhớ mới cho các token, mà thay vào đó nó sửa đổi chuỗi nguồn. Điều này có nghĩa là chuỗi nguồn không thể là **const** (không thể là một con trỏ tới chuỗi cố định). Nó cũng có nghĩa là danh tính của ký tự phân tách bị mất (ví dụ: trong ví dụ trên, các dấu , và ! đã bị xóa khỏi chuỗi nguồn và không thể xác định được ký tự phân tách nào được trùng khớp).

Nếu có nhiều ký tự phân tách liên tiếp trong chuỗi nguồn, chúng sẽ được xem như một ký tự phân tách duy nhất; trong ví dụ, dấu phẩy thứ hai sẽ bị bỏ qua.

strtok() không an toàn khi sử dụng đa luồng hoặc đa ngõ ra do sử dụng **bộ đệm tĩnh** trong quá trình phân tích. Điều này có nghĩa là nếu một hàm gọi strtok(), bất kỳ hàm nào mà nó gọi trong khi sử dụng strtok() cũng không thể sử dụng strtok(), và nó không thể được gọi bởi bất kỳ hàm nào đang sử dụng strtok().

Dưới đây là một ví dụ minh họa cho vấn đề do việc strtok() không đảm bảo tính re-entrant (không đảm bảo khả năng sử dụng lại trong đa luồng). Trong môi trường đa luồng, việc gọi strtok() từ nhiều luồng có thể dẫn đến hành vi không mong muốn và làm mất dữ liệu.

|  |
| --- |
| #include <stdio.h>  #include <string.h>  int main(void)  {      char src[] = "1.2,3.5,4.2";      char \*first = strtok(src, ",");      do      {          char \*part;          /\* Nested calls to strtok do not work as desired \*/          printf("[%s]\n", first);          part = strtok(first, ".");          while (part != NULL)          {                  printf(" [%s]\n", part);                  part = strtok(NULL, ".");          }      } while ((first = strtok(NULL, ",")) != NULL);  } |

**Kết quả:**

|  |
| --- |
| [1.2]  [1]  [2] |

Hoạt động dự kiến là vòng lặp do while bên ngoài sẽ tạo ba mã thông báo bao gồm mỗi chuỗi số thập phân ("1.2", "3.5", "4.2"), đối với mỗi mã mà strtok gọi cho vòng lặp bên trong sẽ chia nó thành các chuỗi riêng biệt chuỗi chữ số ("1", "2", "3", "5", "4", "2").

Tuy nhiên, vì strtok không đăng ký lại nên điều này không xảy ra. Thay vào đó, strtok đầu tiên tạo chính xác mã thông báo "1.2\0" và vòng lặp bên trong tạo chính xác mã thông báo "1" và "2". Nhưng sau đó, strtok trong vòng lặp bên ngoài nằm ở cuối chuỗi được sử dụng bởi vòng lặp bên trong và trả về NULL ngay lập tức. Chuỗi con thứ hai và thứ ba của mảng src hoàn toàn không được phân tích.

Các thư viện C tiêu chuẩn không chứa phiên bản an toàn cho luồng hoặc đăng ký lại nhưng một số phiên bản khác thì có, chẳng hạn như POSIX' strtok\_r. Lưu ý rằng trên MSVC, strtok tương đương, strtok\_s an toàn cho luồng.

C11 có một phần tùy chọn, Phụ lục K, cung cấp phiên bản đăng ký lại và an toàn cho luồng có tên strtok\_s. Bạn có thể kiểm tra tính năng này bằng \_\_STDC\_LIB\_EXT1\_\_. Phần tùy chọn này không được hỗ trợ rộng rãi.

Hàm strtok\_s khác với hàm strtok\_r của POSIX ở chỗ bảo vệ chống lưu trữ bên ngoài chuỗi được mã hóa và bằng cách kiểm tra các giới hạn thời gian chạy. Tuy nhiên, trên các chương trình được viết chính xác, strtok\_s và strtok\_r hoạt động giống nhau.

Sử dụng strtok\_s với ví dụ hiện mang lại phản hồi chính xác, như vậy:

## PHẦN 6.2 STRING LITERALS

Chuỗi hằng (string literals) trong C đại diện cho các mảng ký tự có kết thúc bằng ký tự null và có thời gian sống tĩnh (static-duration). Vì có thời gian sống tĩnh, một chuỗi hằng hoặc một con trỏ trỏ tới cùng một mảng gốc có thể an toàn sử dụng trong nhiều trường hợp mà một con trỏ tới mảng tự động không thể. Ví dụ, việc trả về một chuỗi hằng từ một hàm có hành vi xác định rõ ràng:

|  |
| --- |
| const char \*get\_hello() {      return "Hello, World!"; /\* an toàn \*/  } |

Do lưu trữ tĩnh, các chuỗi hằng không thể bị sửa đổi, và mọi cố gắng thay đổi các phần tử của mảng tương ứng của chuỗi hằng sẽ dẫn đến hành vi không xác định (undefined behavior). Thông thường, một chương trình cố gắng sửa đổi mảng tương ứng của chuỗi hằng sẽ gây ra lỗi hoặc không hoạt động đúng.

|  |
| --- |
| char \*foo = "hello";  foo[0] = 'y'; /\* Hành vi không xác định - KHÔNG TỐT! \*/ |

Để tránh việc sửa đổi không an toàn như vậy, khi một con trỏ trỏ tới một chuỗi hằng hoặc có thể trỏ tới nó, chúng ta nên khai báo con trỏ đó là **const** để tránh vô tình gây ra hành vi không xác định.

|  |
| --- |
| const char \*foo = "hello";  /\* TỐT: không thể sửa đổi chuỗi mà foo trỏ tới \*/ |

Tuy nhiên, một con trỏ trỏ vào hoặc trỏ vào mảng gốc của một chuỗi hằng không có tính đặc biệt về bản chất. Chúng ta có thể tự do sửa đổi giá trị của con trỏ để trỏ tới một địa chỉ khác.

|  |
| --- |
| char \*foo = "hello";  foo = "World!"; /\* OK - chúng ta chỉ đổi địa chỉ mà foo trỏ tới \*/ |

Mặt khác, việc khởi tạo một mảng ký tự bằng cách sử dụng cú pháp giống như chuỗi hằng không làm cho mảng đã khởi tạo có các đặc tính của chuỗi hằng. Khởi tạo chỉ đơn giản xác định độ dài và nội dung ban đầu của mảng. Các phần tử trong mảng này vẫn có thể được sửa đổi nếu chúng không được khai báo là hằng số.

|  |
| --- |
| char foo[] = "hello";  foo[0] = 'y'; /\* OK! \*/ |

## PHẦN 6.3 TÍNH ĐỘ DÀI CHUỖI: STRLEN()

|  |
| --- |
| #include <stdio.h>  #include <stdlib.h>  #include <string.h>  int main(int *argc*, char \*\**argv*)  {      /\* Exit if no second argument is found. \*/      if (*argc* != 2)      {          puts("Argument missing.");          return EXIT\_FAILURE;      }  *size\_t* len = strlen(*argv*[1]);      printf("The length of the second argument is %zu.\\n", len);      return EXIT\_SUCCESS;  } |

Chương trình trên tính độ dài của **đối số thứ hai** mà nó nhận được và lưu kết quả vào biến **len**. Sau đó, nó in ra độ dài đó lên màn hình. Ví dụ, khi chạy với các tham số program\_name "Hello, world!", chương trình sẽ in ra **The length of the second argument is 13**. bởi vì chuỗi Hello, world! có độ dài là 13 ký tự.

Hàm **strlen** đếm tất cả các byte từ đầu chuỗi cho đến ký tự kết thúc **NULL ('\0')**. Vì vậy, nó chỉ có thể sử dụng khi đảm bảo chuỗi kết thúc bằng NULL.

Tuy nhiên, hãy lưu ý rằng nếu chuỗi chứa các ký tự Unicode, hàm **strlen** sẽ không cho bạn biết có bao nhiêu ký tự trong chuỗi (vì một số ký tự có thể dài hơn một byte). Trong trường hợp này, bạn phải tự đếm các ký tự (đơn vị mã) của chuỗi. Ví dụ dưới đây giải thích điều này:

|  |
| --- |
| #include <stdio.h>  #include <stdlib.h>  #include <string.h>  int main(void) {      char asciiString[50] = "Hello world!";      char utf8String[50] = "Γειά σου Κόσμε!"; /\* "Hello World!" in Greek \*/      printf("asciiString co %zu byte trong mang\n", sizeof(asciiString));      printf("utf8String co %zu byte trong mang\n", sizeof(utf8String));      printf("\"%s\" co %zu byte\n", asciiString, strlen(asciiString));      printf("\"%s\" co %zu byte\n", utf8String, strlen(utf8String));      return 0;  } |

**Kết quả**

|  |
| --- |
| asciiString co 50 byte trong mang  utf8String co 50 byte trong mang  "Hello world!" co 12 byte  "╬ô╬╡╬╣╬¼ ╧â╬┐╧à ╬Ü╧î╧â╬╝╬╡!" co 27 byte |

## PHẦN 6.4 GIỚI THIỆU CƠ BẢN VỀ CHUỖI

Trong ngôn ngữ lập trình C, một chuỗi (string) là một chuỗi các ký tự kết thúc bằng ký tự null (‘\0’).

Chúng ta có thể tạo các chuỗi sử dụng chuỗi ký tự (**string literals**), là các chuỗi ký tự được bao quanh bởi dấu ngoặc kép; ví dụ, **"hello world".** Chuỗi ký tự được tự động kết thúc bằng ký tự **null.**

Chúng ta có thể tạo các chuỗi bằng nhiều phương pháp khác nhau. Ví dụ, chúng ta có thể khai báo một con trỏ kiểu char và khởi tạo nó để trỏ đến ký tự đầu tiên của một chuỗi:

|  |
| --- |
| char \*string = "hello world"; |

Khi khởi tạo một con trỏ char bằng một chuỗi hằng như trên, chuỗi thực tế thường được cấp phát trong bộ nhớ chỉ đọc (read-only data); string là một con trỏ trỏ đến phần tử đầu tiên của mảng, tức là ký tự '**h**'.

Vì chuỗi hằng này được cấp phát trong bộ nhớ chỉ đọc, nó không thể thay đổi. Bất kỳ thay đổi nào của nó sẽ dẫn đến hành vi không xác định, do đó tốt hơn hãy thêm từ khóa const để có một thông báo lỗi tại thời gian biên dịch như sau:

|  |
| --- |
| char const \*string = "hello world"; |

Nó có tác dụng tương tự như sau:

|  |
| --- |
| char const string\_arr[] = "hello world"; |

Để tạo một chuỗi có thể sửa đổi, bạn có thể khai báo một mảng ký tự và khởi tạo nội dung của nó bằng chuỗi ký tự, như sau:

|  |
| --- |
| char modifiable\_string[] = "hello world"; |

Điều này tương đương với:

|  |
| --- |
| char modifiable\_string[] = {'h', 'e', 'l', 'l', 'o', ' ', 'w', 'o', 'r', 'l', 'd', '\0'}; |

Vì phiên bản thứ hai sử dụng cách khởi tạo bằng dấu ngoặc nhọn, chuỗi không tự động được kết thúc bằng ký tự null trừ khi ký tự **'\0'** được bao gồm rõ ràng trong mảng ký tự, thường là phần tử cuối cùng của mảng.

**Lưu ý:**

* Từ "không thể sửa đổi" (non-modifiable) chỉ ám chỉ rằng các ký tự trong chuỗi hằng không thể thay đổi, nhưng hãy nhớ rằng con trỏ string có thể được sửa đổi (có thể trỏ tới một vị trí khác hoặc tăng hoặc giảm giá trị).
* Cả hai chuỗi đều có hiệu quả tương tự trong việc không cho phép thay đổi các ký tự của chúng. Lưu ý rằng string là một con trỏ kiểu char và nó là một l-value có thể sửa đổi, nó có thể được tăng giá trị hoặc trỏ đến một vị trí khác trong khi mảng string\_arr là một l-value không thể sửa đổi, nó không thể thay đổi.

## PHẦN 6.5 SAO CHÉP CHUỖI

### 6.5.1 Gán con trỏ không sao chép chuỗi

Trong C, việc sử dụng toán tử = (gán) không sao chép chuỗi. Chuỗi trong C được biểu diễn dưới dạng mảng các ký tự kết thúc bằng ký tự null, vì vậy việc sử dụng toán tử = chỉ lưu địa chỉ (con trỏ) của chuỗi đó.

|  |
| --- |
| #include <stdio.h>  int main(void) {      int a = 10, b;      char c[] = "abc", \*d;      b = a; /\* Integer is copied \*/      a = 20; /\* Modifying a leaves b unchanged - b is a 'deep copy' of a \*/      printf("%d %d\n", a, b); /\* "20 10" will be printed \*/      d = c;      /\* Only copies the address of the string -      there is still only one string stored in memory \*/        c[1] = 'x';      /\* Modifies the original string - d[1] = 'x' will do exactly the same thing \*/      printf("%s %s\n", c, d); /\* "axc axc" will be printed \*/      return 0;  } |

Đoạn mã nguồn trên đã biên dịch thành công vì chúng ta sử dụng con trỏ char \*d thay vì mảng char d[3]. Khi sử dụng mảng char d[3], chương trình sẽ gặp lỗi biên dịch vì trong C, bạn không thể gán trực tiếp cho một mảng.

|  |
| --- |
| #include <stdio.h>  int main(void) {      char a[] = "abc";      char b[8];      b = a; /\* compile error \*/      printf("%s\n", b);      return 0;  } |

### 6.5.2 Sao chép chuỗi dùng các hàm tiêu chuẩn

#### 6.5.2.1 strcpy()

Để sao chép chuỗi, ta có thể sử dụng hàm strcpy() trong thư viện string.h. Trước khi sao chép, ta cần phải cấp phát đủ không gian cho chuỗi đích.

|  |
| --- |
| #include <stdio.h>  #include <string.h>  int main(void) {      char a[] = "abc";      char b[8];      strcpy(b, a); /\* think "b special equals a" \*/      printf("%s\n", b); /\* "abc" will be printed \*/      return 0;  } |

#### 6.5.2.2 snprintf()

Để tránh tràn bộ đệm, snprintf() có thể được sử dụng. Đây không phải là giải pháp tốt nhất về mặt hiệu năng vì nó phải phân tích cú pháp chuỗi mẫu, nhưng đây là chức năng an toàn giới hạn bộ đệm duy nhất để sao chép các chuỗi có sẵn trong thư viện chuẩn, có thể được sử dụng mà không cần thực hiện thêm bất kỳ bước nào.

Hàm snprintf() hữu ích khi ta muốn sao chép một chuỗi vào một bộ đệm cố định với kích thước xác định mà không muốn gây ra lỗi tràn bộ đệm (buffer overflow). Thay vì sao chép toàn bộ nội dung của chuỗi như hàm strcpy(), hàm snprintf() cho phép bạn chỉ định kích thước tối đa cho bộ đệm đích.

|  |
| --- |
| #include <stdio.h>  #include <string.h>  int main(void) {      char a[] = "012345678901234567890";      char b[8];  #if 0   strcpy(b, a); /\* causes buffer overrun (undefined behavior), so do not execute this here! \*/  #endif      snprintf(b, sizeof(b), "%s", a); /\* does not cause buffer overrun \*/      printf("%s\n", b); /\* "0123456" will be printed \*/      return 0;  } |

#### 6.5.2.3 strncat()

Một lựa chọn thứ hai, với hiệu suất tốt hơn, là sử dụng hàm strncat() (phiên bản kiểm tra tràn bộ đệm của strcat()) - nó nhận thêm một đối số thứ ba để xác định số byte tối đa cần sao chép.

|  |
| --- |
| char dest[32];  dest[0] = '\0'; // Đảm bảo dest là chuỗi rỗng ban đầu  strncat(*dest*, *source*, sizeof(dest) - 1);  // Sao chép tối đa (sizeof(dest) - 1) phần tử từ source vào dest  // Sau đó, thêm ký tự null '\0' vào cuối chuỗi dest |

Lưu ý rằng trong công thức trên, chúng ta sử dụng sizeof(dest) - 1. Điều này quan trọng vì hàm strncat() luôn thêm một ký tự null ('\0') vào cuối chuỗi, nhưng không tính ký tự này trong kích thước của chuỗi (điều này có thể gây nhầm lẫn và tràn bộ đệm).

Một ví dụ khác là khi ghép chuỗi sau một chuỗi không rỗng:

|  |
| --- |
| char dst[24] = "Clownfish: ";  char src[] = "Marvin and Nemo";  *size\_t* len = strlen(dst);  strncat(*dst*, *src*, sizeof(dst) - len - 1);  printf("%zu: [%s]\n", strlen(*dst*), *dst*); |

Kết quả khi in ra màn hình sẽ là:

|  |
| --- |
| 23: [Clownfish: Marvin and N] |

Lưu ý rằng kích thước được chỉ định là độ dài còn lại của mảng đích (không tính ký tự null cuối cùng). Điều này có thể gây ra vấn đề ghi đè lớn hơn. Để xác định đúng độ dài cần sao chép, bạn nên chỉ định địa chỉ của ký tự null cuối cùng trong nội dung hiện có của chuỗi đích, giúp strncat() không phải quét lại chuỗi đích trước khi bắt đầu sao chép.

|  |
| --- |
| strcpy(*dst*, "Clownfish: ");  assert(len < sizeof(dst) - 1);  strncat(dst + len, src, sizeof(dst) - len - 1);  printf("%zu: [%s]\n", strlen(*dst*), *dst*); |

Điều này tạo ra cùng kết quả như trước, nhưng strncat() không cần quét lại nội dung hiện có của dst trước khi bắt đầu sao chép.

#### 6.5.2.4 strncpy()

Hàm strncpy() là lựa chọn cuối cùng. Mặc dù có thể nghĩ rằng nó nên đứng đầu, nhưng nó lại là một hàm gây nhầm lẫn với hai vấn đề chính:

* Nếu sao chép bằng strncpy() vượt quá giới hạn của bộ đệm, một ký tự null cuối cùng sẽ không được sao chép vào đích.
* strncpy() luôn điền đầy đủ nội dung của đích, bằng các ký tự null nếu cần thiết.

Hiện tại, việc triển khai kỳ quặc này có lịch sử và ban đầu được dùng để xử lý tên tập tin UNIX.)

Cách duy nhất để sử dụng strncpy() đúng đắn là tự đảm bảo kết thúc chuỗi bằng ký tự null:

|  |
| --- |
| strncpy(*b*, *a*, sizeof(b)); /\* tham số thứ ba là kích thước của bộ đệm đích \*/  b[sizeof(b)/sizeof(\*b) - 1] = '\0'; /\* kết thúc chuỗi \*/  printf("%s\n", b); /\* "0123456" sẽ được in ra \*/ |

Tuy nhiên, thậm chí khi bạn có một bộ đệm lớn, việc sử dụng strncpy() trở nên rất không hiệu quả do cần phải điền ký tự null thừa.

## PHẦN 6.6 LẶP QUA CÁC KÍ TỰ TRONG MỘT CHUỖI

Nếu chúng ta biết độ dài của chuỗi, chúng ta có thể sử dụng vòng lặp for để lặp lại các ký tự của chuỗi:

|  |
| --- |
| char \* string = "hello world"; /\* This 11 chars long, excluding the 0-terminator. \*/  *size\_t* i = 0;  for (; i < 11; i++) {      printf("%c\n", string[i]); /\* Print each character of the string. \*/  } |

Ngoài ra, chúng ta có thể sử dụng hàm tiêu chuẩn strlen() để lấy độ dài của một chuỗi nếu chúng ta không biết chuỗi đó là gì:

|  |
| --- |
| size\_t length = strlen(string);  *size\_t* i = 0;  for (; i < length; i++) {      printf("%c\n", string[i]); /\* Print each character of the string. \*/  } |

Cuối cùng, chúng ta có thể tận dụng lợi thế của thực tế là các chuỗi trong C được đảm bảo là kết thúc null (điều mà chúng ta đã làm khi chuyển nó sang strlen() trong ví dụ trước ;-)). Chúng tôi có thể lặp lại mảng bất kể kích thước của nó và dừng lặp lại khi chúng tôi đạt đến ký tự null:

|  |
| --- |
| size\_t i = 0;  while (string[i] != '\0') { /\* Stop looping when we reach the null-character. \*/      printf("%c\n", string[i]); /\* Print each character of the string. \*/      i++;  } |

## PHẦN 6.7 TẠO MẢNG CHUỖI

Một mảng chuỗi trong C có thể có ý nghĩa như sau:

* Một mảng có các phần tử là các con trỏ char \*.
* Một mảng có các phần tử là các mảng ký tự.

Chúng ta có thể tạo một mảng các con trỏ ký tự như sau:

|  |
| --- |
| char \* string\_array[] = {      "foo",      "bar",      "baz"  }; |

Hãy nhớ rằng khi gán các chuỗi chữ cho con trỏ **char \***, các chuỗi đó thực sự được lưu trữ trong bộ nhớ chỉ đọc (**read-only**). Tuy nhiên, mảng string\_array được lưu trữ trong bộ nhớ có thể đọc và viết (**read/write**). Điều này có nghĩa là chúng ta có thể chỉnh sửa các con trỏ trong mảng, nhưng không thể chỉnh sửa các chuỗi mà chúng trỏ tới.

Trong C, tham số argv của hàm main (mảng các tham số dòng lệnh được truyền khi chạy chương trình) là một mảng các con trỏ char \*: char \*argv[].

Chúng ta cũng có thể tạo mảng các mảng ký tự. Vì chuỗi là mảng ký tự, một mảng chuỗi đơn giản là một mảng có các phần tử là các mảng ký tự:

|  |
| --- |
| char modifiable\_string\_array\_literals[][4] = {      "foo",      "bar",      "baz"  }; |

Điều này tương đương với:

|  |
| --- |
| char modifiable\_string\_array[][4] = {      {'f', 'o', 'o', '\0'},      {'b', 'a', 'r', '\0'},      {'b', 'a', 'z', '\0'}  }; |

Lưu ý rằng chúng ta chỉ định 4 là kích thước của chiều thứ hai của mảng; mỗi chuỗi trong mảng của chúng ta thực sự có kích thước 4 byte vì chúng ta phải bao gồm ký tự kết thúc chuỗi **(null-terminating character).**

## PHẦN 6.8 CHUYỂN ĐỔI CHUỖI THÀNH SỐ: ATOI(), ATOF() (CẨN THẬN KHÔNG NÊN SỬ DỤNG)

Cảnh báo: Các hàm atoi, atol, atoll và atof inherently là không an toàn, bởi vì: Nếu giá trị của kết quả không thể biểu diễn, hành vi của chúng không xác định. (7.20.1p1)

|  |
| --- |
| #include <stdio.h>  #include <stdlib.h>  int main(int *argc*, char\*\* *argv*)  {      int val;      if (*argc* < 2)      {          printf("Usage: %s <integer>\n", *argv*[0]);          return 0;      }      val = atoi(*argv*[1]);      printf("String value = %s, Int value = %d\n", *argv*[1], val);      return 0;  } |

Khi chuỗi cần chuyển đổi là một số nguyên hợp lệ và nằm trong phạm vi của kiểu dữ liệu int, thì hàm atoi() hoạt động đúng:

|  |
| --- |
| $ ./atoi 100  String value = 100, Int value = 100  $ ./atoi 200  String value = 200, Int value = 200 |

Tuy nhiên, nếu chuỗi bắt đầu bằng một số, tiếp theo là ký tự khác, thì chỉ số đầu tiên của chuỗi đó sẽ được chuyển đổi:

|  |
| --- |
| $ ./atoi 0x200  0  $ ./atoi 0123x300  123 |

Nếu chuỗi không phải là một số hợp lệ hoặc không nằm trong phạm vi kiểu dữ liệu int, thì hành vi của hàm atoi() sẽ là không xác định, có thể gây ra lỗi hoặc dẫn đến các hậu quả không mong muốn. Ví dụ:

|  |
| --- |
| $ ./atoi hello  Formatting the hard disk... |

Do sự mơ hồ ở trên và hành vi không xác định này, họ các chức năng atoi không bao giờ nên được sử dụng.

Để chuyển đổi thành long int, hãy sử dụng strtol() thay vì atol().

Để chuyển đổi thành double, hãy sử dụng strtod() thay vì atof().

Để chuyển đổi thành long long int, hãy sử dụng strtoll() thay vì atoll().

## PHẦN 6.9 ĐỌC GHI DỮ LIỆU DẠNG CHUỖI

Ghi dữ liệu đã định dạng vào chuỗi

|  |
| --- |
| int sprintf ( char \* *str*, const char \* *format*, ... ); |

sử dụng hàm **sprintf** để ghi dữ liệu float vào chuỗi.

|  |
| --- |
| #include <stdio.h>  int main ()  {      char buffer [50];      double PI = 3.1415926;      sprintf (buffer, "PI = %.7f", PI);      printf ("%s\n",buffer);      return 0;  } |

Kết quả

|  |
| --- |
| PI = 3.1415926 |

Đọc dữ liệu được định dạng từ chuỗi

|  |
| --- |
| int sscanf ( const char \* *s*, const char \* *format*, ...); |

sử dụng hàm **sscanf** để phân tích dữ liệu được định dạng.

|  |
| --- |
| #include <stdio.h>  int main ()  {      char sentence []="date : 06-06-2012";      char str [50];      int year;      int month;      int day;      sscanf (sentence,"%s : %2d-%2d-%4d", str, &day, &month, &year);      printf ("%s -> %02d-%02d-%4d\n",str, day, month, year);      return 0;  } |

Kết quả:

|  |
| --- |
| date -> 06-06-2012 |

## PHẦN 6.10 TÌM VỊ TRÍ XUẤT HIỆN ĐẦU TIÊN/ CUỐI CÙNG CỦA MỘT KÍ TỰ CỤ THỂ: STRCHR(), STRRCHR()

Các hàm **strchr** và **strrchr** tìm một ký tự trong một chuỗi, ký tự đó nằm trong một mảng ký tự kết thúc bằng NULL. **strchr** trả về một con trỏ tới lần xuất hiện đầu tiên và **strrchr** tới lần xuất hiện cuối cùng.

|  |
| --- |
| #include <stdio.h>  #include <stdlib.h>  #include <string.h>  int main(int *argc*, char \**argv*[]) // Thêm khai báo biến argc và argv ở đây  {      char toSearchFor = 'A';      /\* Exit if no second argument is found. \*/      if (*argc* != 2)      {          printf("Argument missing.\n");          return EXIT\_FAILURE;      }      {          char \*firstOcc = strchr(*argv*[1], toSearchFor);          if (firstOcc != NULL)          {              printf("First position of %c in %s is %td.\n",                     toSearchFor, *argv*[1], firstOcc - *argv*[1]);              /\* A pointer difference's result is a signed integer and uses the length modifier 't'. \*/          }          else          {              printf("%c is not in %s.\n", toSearchFor, *argv*[1]);          }      }      {          char \*lastOcc = strrchr(*argv*[1], toSearchFor);          if (lastOcc != NULL)          {              printf("Last position of %c in %s is %td.\n",                     toSearchFor, *argv*[1], lastOcc - *argv*[1]);          }      }      return EXIT\_SUCCESS;  } |

Kết quả đầu ra (sau khi đã tạo một pos có tên thực thi):

|  |
| --- |
| $ ./pos AAAAAAA  First position of A in AAAAAAA is 0.  Last position of A in AAAAAAA is 6.  $ ./pos BAbbbbbAccccAAAAzzz  First position of A in BAbbbbbAccccAAAAzzz is 1.  Last position of A in BAbbbbbAccccAAAAzzz is 15.  $ ./pos qwerty  A is not in qwerty. |

Một cách sử dụng phổ biến cho strrchr là trích xuất tên tệp từ một đường dẫn. Ví dụ: để giải nén myfile.txt từ C:\Users\eak\myfile.txt:

|  |
| --- |
| char \*getFileName(const char \**path*)  {      char \*pend;      if ((pend = strrchr(*path*, '\')) != NULL)          return pend + 1;      return NULL;  } |

## PHẦN 6.11 SAO CHÉP VÀ NỐI CHUỖI: STRCPY(), STRCAT()

|  |
| --- |
| #include <stdio.h>  #include <string.h>  int main(void) {      /\* Khai báo một mảng có đủ độ dài để chứa các ký tự và ký tự null kết thúc chuỗi ('\0'). \*/      char mystring[10];      /\* Sao chép chuỗi "foo" vào `mystring` cho đến khi gặp ký tự null. \*/      strcpy(mystring, "foo");      printf("%s\n", mystring);      /\* Tại thời điểm này, chúng ta đã sử dụng 4 ký tự trong `mystring`, bao gồm 3 ký tự của "foo" và ký tự null kết thúc. \*/      /\* Nối chuỗi "bar" vào `mystring`. \*/      strcat(mystring, "bar");      printf("%s\n", mystring);      /\* Bây giờ chúng ta đã sử dụng 7 ký tự trong `mystring`: "foo" dùng 3 ký tự, "bar" dùng 3 ký tự và ký tự null kết thúc. \*/      /\* Sao chép chuỗi "bar" vào `mystring`, ghi đè nội dung cũ. \*/      strcpy(mystring, "bar");      printf("%s\n", mystring);      return 0;  } |

Kết quả:

|  |
| --- |
| foo  foobar  bar |

Khi bạn nối vào, sao chép từ hoặc nối từ một chuỗi hiện có, hãy đảm bảo rằng chuỗi này đã được kết thúc bằng ký tự null trước khi thực hiện thao tác. Trường hợp của bạn ví dụ, chuỗi từ hằng số (string literals) như "foo" đã được biên dịch sẽ luôn được kết thúc bằng ký tự null bởi trình biên dịch, vì vậy bạn có thể yên tâm sử dụng chúng trong các hàm xử lý chuỗi.

## PHẦN 6.12 SO SÁNH: STRCMP(), STRNCMP(), STRCASECMP(), STRNCASECMP()

Các hàm strcase\*-không phải là Tiêu chuẩn C, mà là phần mở rộng POSIX.

strcmp(): Hàm này so sánh hai chuỗi ký tự kết thúc bằng null (null-terminated) theo thứ tự từ điển. Hàm trả về một giá trị số nguyên:

* Nếu chuỗi thứ nhất xuất hiện trước chuỗi thứ hai trong thứ tự từ điển, hàm trả về một giá trị âm.
* Nếu hai chuỗi bằng nhau, hàm trả về giá trị 0.
* Nếu chuỗi thứ nhất xuất hiện sau chuỗi thứ hai trong thứ tự từ điển, hàm trả về một giá trị dương.

|  |
| --- |
| #include <stdio.h>  #include <string.h>  void compare(char const \**lhs*, char const \**rhs*)  {      int result = strcmp(*lhs*, *rhs*); // compute comparison once      if (result < 0) {          printf("%s comes before %s\n", *lhs*, *rhs*);      } else if (result == 0) {          printf("%s equals %s\n", *lhs*, *rhs*);      } else { // last case: result > 0          printf("%s comes after %s\n", *lhs*, *rhs*);      }  }  int main(void)  {      compare("BBB", "BBB");      compare("BBB", "CCCCC");      compare("BBB", "AAAAAA");      return 0;  } |

Kết quả:

|  |
| --- |
| BBB equals BBB  BBB comes before CCCCC  BBB comes after AAAAAA |

Giống như hàm **strcmp**, hàm **strcasecmp** cũng so sánh các đối số của nó theo từ điển sau khi dịch từng ký tự sang chữ thường tương ứng của nó:

|  |
| --- |
| #include <stdio.h>  #include <string.h>  void compare(char const \**lhs*, char const \**rhs*)  {      int result = strcasecmp(*lhs*, *rhs*); // compute case-insensitive comparison once      if (result < 0) {          printf("%s comes before %s\n", *lhs*, *rhs*);      } else if (result == 0) {          printf("%s equals %s\n", *lhs*, *rhs*);      } else { // last case: result > 0          printf("%s comes after %s\n", *lhs*, *rhs*);      }  }  int main(void)  {      compare("BBB", "bBB");      compare("BBB", "ccCCC");      compare("BBB", "aaaaaa");      return 0;  } |

Kết quả:

|  |
| --- |
| BBB equals bBB  BBB comes before ccCCC  BBB comes after aaaaaa |

**strncmp** và **strncasecmp** so sánh tối đa n ký tự. Hàm này nhận thêm một tham số thứ ba là số lượng ký tự để so sánh:

|  |
| --- |
| #include <stdio.h>  #include <string.h>  void compare(char const \**lhs*, char const \**rhs*, int *n*)  {      int result = strncmp(*lhs*, *rhs*, *n*); // compute comparison once      if (result < 0) {          printf("%s comes before %s\n", *lhs*, *rhs*);      } else if (result == 0) {          printf("%s equals %s\n", *lhs*, *rhs*);      } else { // last case: result > 0          printf("%s comes after %s\n", *lhs*, *rhs*);      }  }  int main(void)  {      compare("BBB", "Bb", 1);      compare("BBB", "Bb", 2);      compare("BBB", "Bb", 3);      return 0;  } |

Kết quả:

|  |
| --- |
| BBB equals Bb  BBB comes before Bb  BBB comes before Bb |

## PHẦN 6.13 CHUYỂN ĐỔI CHUỖI THÀNH SỐ MỘT CÁCH AN TOÀN: HÀM STRTOX

Kể từ C99, thư viện C có một tập hợp các hàm chuyển đổi an toàn diễn giải một chuỗi thành một số. Tên của chúng có dạng strtoX, trong đó X là một trong số l, ul, d, v.v. để xác định loại mục tiêu của chuyển đổi.

strtol: Chuyển đổi chuỗi sang số nguyên dạng long (32-bit hoặc 64-bit tùy thuộc vào hệ điều hành và kiến trúc).

strtoul: Chuyển đổi chuỗi sang số nguyên không dấu dạng unsigned long.

strtoll: Chuyển đổi chuỗi sang số nguyên dạng long long (64-bit).

strtoull: Chuyển đổi chuỗi sang số nguyên không dấu dạng unsigned long long.

strtod: Chuyển đổi chuỗi sang số thực dạng double.

|  |
| --- |
| double strtod(char const\* *p*, char\*\* *endptr*);  long double strtold(char const\* *p*, char\*\* *endptr*); |

Chúng cung cấp khả năng kiểm tra xem chuyển đổi có bị tràn hay thiếu không:

|  |
| --- |
| double ret = strtod(argv[1], 0); /\* attempt conversion \*/  /\* check the conversion result. \*/  if ((ret == HUGE\_VAL || ret == -HUGE\_VAL) && errno == ERANGE)      return; /\* numeric overflow in in string \*/  else if (ret == HUGE\_VAL && errno == ERANGE)      return; /\* numeric underflow in in string \*/  /\* At this point we know that everything went fine so ret may be used \*/ |

Nếu chuỗi không chứa số nào, hàm **strtod** trên sẽ trả về giá trị 0.0.

Tuy nhiên, nếu không muốn nhận giá trị mặc định này, ta có thể sử dụng tham số bổ sung **endptr**. Tham số này là một con trỏ tới con trỏ, và nó sẽ được trỏ tới cuối số đã được phát hiện trong chuỗi. Nếu tham số này được đặt thành 0 hoặc NULL, nó sẽ được bỏ qua.

Sử dụng tham số **endptr**, ta có thể kiểm tra xem quá trình chuyển đổi có thành công hay không và xác định vị trí kết thúc của số trong chuỗi:

|  |
| --- |
| char \*check = 0;  double ret = strtod(argv[1], &*check*); /\* thử chuyển đổi \*/  if (argv[1] == check)      return; /\* Không phát hiện số nào trong chuỗi \*/  else if ((ret == HUGE\_VAL || ret == -HUGE\_VAL) && errno == ERANGE)      return; /\* Tràn số khi chuyển đổi \*/  else if (ret == HUGE\_VAL && errno == ERANGE)      return; /\* Số quá nhỏ khi chuyển đổi \*/  /\* Ở đây, ta biết rằng mọi thứ đã diễn ra suôn sẻ, giá trị ret có thể sử dụng \*/ |

Có các hàm tương tự để chuyển đổi sang các loại số nguyên rộng hơn:

|  |
| --- |
| long strtol(char const\* *p*, char\*\* *endptr*, int *nbase*);  long long strtoll(char const\* *p*, char\*\* *endptr*, int *nbase*);  unsigned long strtoul(char const\* *p*, char\*\* *endptr*, int *nbase*);  unsigned long long strtoull(char const\* *p*, char\*\* *endptr*, int *nbase*); |

Các hàm này có tham số thứ ba nbase giữ cơ sở số trong đó số được viết.

|  |
| --- |
| long a = strtol("101", 0, 2 ); /\* a = 5L \*/  long b = strtol("101", 0, 8 ); /\* b = 65L \*/  long c = strtol("101", 0, 10); /\* c = 101L \*/  long d = strtol("101", 0, 16); /\* d = 257L \*/  long e = strtol("101", 0, 0 ); /\* e = 101L \*/  long f = strtol("0101", 0, 0 ); /\* f = 65L \*/  long g = strtol("0x101", 0, 0 ); /\* g = 257L \*/ |

Giá trị 0 cho nbase có nghĩa là chuỗi sẽ được hiểu như các số trong chương trình C: tiền tố 0x tương ứng với số thập lục phân, tiền tố 0 tương ứng với số bát phân, và tất cả các số khác được hiểu là số thập phân.

Do đó, cách thiết thực nhất để diễn giải một đối số dòng lệnh dưới dạng số sẽ là

|  |
| --- |
| int main(int *argc*, char\* *argv*[] {      if (*argc* < 1)          return EXIT\_FAILURE; /\* No number given. \*/          /\* use strtoull because size\_t may be wide \*/  *size\_t* mySize = strtoull(*argv*[1], 0, 0);          /\* then check conversion results. \*/          ...          return EXIT\_SUCCESS;  } |

Điều này cho phép chương trình được gọi với một tham số dòng lệnh ở hệ thống bát phân, thập phân hoặc thập lục phân.

## PHẦN 6.14 STRSPN VÀ STRCSPN

Cho trước một chuỗi, strspn tính toán độ dài của chuỗi con ban đầu (span) chỉ bao gồm một danh sách ký tự cụ thể. strcspn cũng tương tự, ngoại trừ nó tính toán độ dài của chuỗi con ban đầu bao gồm bất kỳ ký tự nào ngoại trừ những ký tự được liệt kê:

|  |
| --- |
| /\*  Provided a string of "tokens" delimited by "separators", print the tokens along  with the token separators that get skipped.  \*/  #include <stdio.h>  #include <string.h>  int main(void)  {      const char sepchars[] = ",.;!?";      char foo[] = ";ball call,.fall gall hall!?.,";      char \*s;      int n;      for (s = foo; \*s != 0; /\*empty\*/) {          /\* Get the number of token separator characters. \*/          n = (int)strspn(s, sepchars);          if (n > 0)              printf("skipping separators: << %.\*s >> (length=%d)\n", n, s, n);          /\* Actually skip the separators now. \*/          s += n;          /\* Get the number of token (non-separator) characters. \*/          n = (int)strcspn(s, sepchars);          if (n > 0)              printf("token found: << %.\*s >> (length=%d)\n", n, s, n);          /\* Skip the token now. \*/          s += n;      }      printf("== token list exhausted ==\n");      return 0;  } |

Các hàm tương tự sử dụng chuỗi ký tự rộng là wcsspn và wcscspn; chúng được sử dụng theo cùng một cách.

# CHỮ SỐ, KÝ TỰ VÀ CHUỖI

## PHẦN 7.1 FLOATING POINT LITERALS( CHỮ SỐ DẤU CHẤM ĐỘNG)

Chữ dấu chấm động được sử dụng để biểu diễn các số thực có dấu. Các hậu tố sau đây có thể được sử dụng để chỉ định loại chữ:

|  |  |  |
| --- | --- | --- |
| Hậu tố | Kiểu dữ liệu | Ví dụ |
| Không có | double | 3.1415926 -3E6 |
| f, F | float | 3.1415926f 2.1E-6F |
| l, L | long double | 3.1415926L 1E126L |

Để sử dụng các hậu tố này, hằng số phải là một hằng số số thực dấu chấm động. Ví dụ, 3f là không hợp lệ, vì 3 là một hằng số số nguyên, trong khi 3.f hoặc 3.0f là đúng. Đối với kiểu long double, khuyến nghị luôn sử dụng chữ hoa L để cải thiện tính đọc hiểu.

## PHẦN 7.2 STRING LITERALS( CHUỖI KÍ TỰ CHỮ)

Chuỗi ký tự chữ được sử dụng để xác định các mảng ký tự. Chúng là các chuỗi ký tự được bao quanh bởi dấu ngoặc kép (ví dụ: "abcd") và có kiểu dữ liệu là char\*.

Tiền tố L làm cho chữ trở thành một mảng ký tự rộng, thuộc loại wchar\_t\*. Ví dụ: L"abcd".

Kể từ phiên bản C11, có các tiền tố mã hóa khác, tương tự như tiền tố L:

|  |  |  |
| --- | --- | --- |
| Tiền tố | Loại cơ sở | Mã hóa |
| Không có | char | tùy thuộc vào nền tảng |
| L | wchar\_t | tùy thuộc vào nền tảng |
| u8 | char | UTF-8 |
| u | char16\_t | Thường là UTF-16 |
| U | char32\_t | Thường là UTF-32 |

Đối với hai tiền tố cuối cùng, ta có thể kiểm tra với các macro kiểm tra tính năng xem liệu mã hóa có phải là mã hóa UTF tương ứng hay không.

## PHẦN 7.3 CHARACTER LITERALS(KÝ TỰ)

Hằng ký tự là một loại đặc biệt của hằng số số nguyên được sử dụng để biểu diễn một ký tự duy nhất. Chúng được bao quanh bởi dấu ngoặc đơn, ví dụ: 'a' và có kiểu dữ liệu là int. Giá trị của hằng số là một giá trị số nguyên tương ứng với bộ ký tự của máy tính. Chúng không cho phép tiền tố.

Tiền tố L trước một hằng ký tự biến nó thành một ký tự rộng của kiểu wchar\_t. Tương tự, từ phiên bản C11, các tiền tố u và U biến chúng thành ký tự rộng của kiểu char16\_t và char32\_t.

Khi muốn biểu diễn một số ký tự đặc biệt như ký tự không in, ta sử dụng các chuỗi escape. Các chuỗi escape sử dụng một chuỗi ký tự được dịch thành một ký tự khác. Tất cả các chuỗi escape đều bao gồm hai hoặc nhiều ký tự, ký tự đầu tiên trong chuỗi là dấu gạch chéo . Các ký tự ngay sau dấu gạch chéo xác định ký tự mà chuỗi escape được hiểu là gì.

Các chuỗi escape được sử dụng để biểu diễn các ký tự như sau:

|  |  |
| --- | --- |
| \b | Dấu phẩy ngược |
| \f | Lùi trang |
| \n | Xuống dòng |
| \r | Cú pháp |
| \t | Tab ngang |
| \v | Tab dọc |
| \\ | Dấu gạch chéo ngược |
| \’ | Dấu nháy đơn |
| \” | Dấu nháy kép |
| \? | Dấu chấm hỏi |
| \nnn | Giá trị bát phân |
| \xnn | Giá trị thập lục phân |
| \a | Ký tự thông báo (còi) |
| \unnnn | Tên ký tự toàn cầu (universal character name) có 4 chữ số thập lục phân |
| \Unnnnnnnn | Tên ký tự toàn cầu có 8 chữ số thập lục phân |

Một tên ký tự toàn cầu là một mã Unicode. Một tên ký tự toàn cầu có thể ánh xạ thành nhiều ký tự. Các chữ số n được hiểu là các chữ số thập lục phân. Tùy thuộc vào mã hóa UTF đang sử dụng, một chuỗi tên ký tự toàn cầu có thể cho ra một mã Unicode bao gồm nhiều ký tự, thay vì chỉ một ký tự duy nhất.

Khi sử dụng chuỗi escape xuống dòng trong I/O chế độ văn bản, nó sẽ được chuyển đổi thành byte xuống dòng cụ thể của hệ điều hành.

Chuỗi escape dấu chấm hỏi được sử dụng để tránh trigraphs. Ví dụ, ??/ sẽ được biên dịch thành trigraph đại diện cho ký tự gạch chéo '', nhưng sử dụng ??/ sẽ cho ra chuỗi "??/". Có thể có một, hai hoặc ba số octal n trong chuỗi escape giá trị bát phân.

## PHẦN 7.4 INTEGER LITERALS(SỐ NGUYÊN)

Hằng số nguyên được sử dụng để cung cấp các giá trị nguyên. Có ba cơ số số học được hỗ trợ, được chỉ định bằng tiền tố:

|  |  |  |
| --- | --- | --- |
| Cơ số | Tiền tố | Ví dụ |
| Thập phân | Không có | 5 |
| Bát phân | 0 | 0345 |
| Thập lục phân | 0x hoặc 0X | 0x12AB, 0X12AB, 0x12ab, 0x12Ab |

Lưu ý rằng viết không bao gồm bất kỳ dấu hiệu nào, do đó hằng số nguyên luôn luôn dương. Một số như -1 được xem như là biểu thức có một hằng số nguyên (1) được đảo dấu bằng dấu -.

Kiểu dữ liệu của hằng số nguyên thập phân là kiểu dữ liệu đầu tiên có thể chứa giá trị từ int và long. Kể từ C99, kiểu long long cũng được hỗ trợ cho các hằng số rất lớn.

Kiểu dữ liệu của hằng số nguyên bát phân hoặc thập lục phân là kiểu dữ liệu đầu tiên có thể chứa giá trị từ int, unsigned, long và unsigned long. Kể từ C99, kiểu long long và unsigned long long cũng được hỗ trợ cho các hằng số rất lớn.

Sử dụng các hậu tố khác nhau, kiểu mặc định của hằng số có thể được thay đổi.

|  |  |
| --- | --- |
| Hậu tố | Giải thích |
| L, l | long int |
| LL, ll( kể tử C99) | long long int |
| U,u | unsigned |

Các hậu tố U và L/LL có thể được kết hợp trong bất kỳ thứ tự và chữ hoa nào. Lỗi nếu có hậu tố bị trùng lặp (ví dụ cung cấp hai hậu tố U) ngay cả khi chúng có chữ hoa khác nhau.