

Bài tập dài số 4

Bài tập dài này yêu cầu thiết kế và xây dựng phần mềm chuyển đổi file dữ liệu sang dạng Morse code và ngược lại.

1. Giới thiệu

Điện tín là phương tiện đầu tiên giúp con người truyền dữ liệu qua khoảng cách lớn. Các thông tin được mã hóa sử dụng Morse code, tương ứng mỗi ký tự alphabet với một mã xác định gồm tập hợp của các dấu . và - . Phương pháp này ứng dụng hiệu ứng âm thanh, với dấu . ứng với âm ngắn và dấu - ứng với âm dài. Người ta còn dùng phương pháp này cho hệ thống gửi tín hiệu qua đường ánh sáng, hoặc dùng cờ hiệu.

Phương pháp mã hóa chuẩn Morse sẽ giải mã mỗi ký hiệu là một tập từ 1 tới 5 dấu . -. Các ký tự đặc biệt sẽ được hiểu là tập của 2, tới 3 ký tự và tương ứng sẽ bao gồm từ 6 dấu . hoặc -. Ví dụ, dấu phẩy , sẽ được hiểu là MIM và sẽ được mã là (--...--). Tám dấu chấm liên tục (.....) sẽ hiểu là lỗi và xóa đi ký tự vừa đánh vào trước.

2. Yêu cầu

Chương trình sẽ mở một file vào và tìm hiểu xem file đấy là file dữ liệu hay file Morse. Sau đó chương trình sẽ chuyển đổi dữ liệu sang dạng tương ứng và kết quả lưu sang file đầu ra xác định

3. Yêu cầu kỹ thuật

Chương trình sẽ bao gồm các chức năng sau:

- Chương trình sẽ chạy từ dòng lệnh (command – line), với tên file chạy là morse, sau đấy tên của 2 file: file đầu vào và file dữ liệu đầu ra.
- Chương trình sẽ kiểm tra dạng của file đầu vào và chuyển đổi dữ liệu sang dạng khác và gửi ra file đầu ra.
- Khi kiểm tra file đầu vào, nếu file chỉ chứa các ký tự (.), (-), (/) và dấu cách (whitespace), thì đó là file Morse, còn lại nếu file có chứa bất cứ ký tự nào khác thì đấy sẽ là file text.
- Các ký tự trong Morse code là cách nhau bởi dấu cách (space)

- Các từ trong Morse code thì cách nhau bởi ký tự /
- Các từ trong file text thì cách nhau bởi dấu cách
- File text chỉ chứa các ký tự thường (ví dụ A tương đương với a). File đầu ra cũng chỉ hiển thị các ký tự thường.
- Các dòng kết thúc bởi dấu \n, và các từ khi in ra không được tràn lên 2 dòng.
- Một dòng đầu vào có thể có chiều dài không giới hạn. Và mỗi dòng đầu vào cần được dịch ra một dòng đầu ra tương ứng.

Mã lỗi Morse code (.....) , sẽ được chuyển sang thành ký tự #. Nếu nó xuất hiện ở giữa một từ, thì ký tự này sẽ được chèn vào từ cần chuyển đổi (ví dụ plai#n-text). Vì thế mà mã lỗi của Morse code sẽ được chuyển đổi thành ký tự bình thường, giống như các mã khác. Tuy nhiên cho file text, ký tự # sẽ được hiểu là lỗi.

Xử lý lỗi

1. Usage: morse <input file> <output file>

Nếu một file mà không mở được thì sẽ có thông báo
Error: FILENAME could not be opened.
với tên file là file không mở được

Nếu tên file đầu ra đã tồn tại thì chương trình sẽ hỏi xem có ghi đè lên file đó không ?

Warning: FILENAME already exists. Do you wish to overwrite (y,n)?
Nếu người dùng ấn y thì ghi đè, nếu không thì kết thúc.

• I Nếu file text có những ký tự mà không có ký tự tương đương trong bảng Morse, thì ký tự đó sẽ bỏ qua và thông báo sau sẽ gửi lên màn hình
Error: Unrecognised character C on line XX.
với c là ký tự mà không hiển thị và xx là dòng mà ký tự đó xuất hiện (tính từ dòng thứ 1)

• I Nếu file Morse code chứa những ký tự mà không có ký tự chữ cái tương đương, thì bạn in ra ký tự * ở file đầu ra và hiện thị lên màn hình thông báo lỗi
Error: Invalid Morse code CODE on line XX.
với CODE là mã Morse mà bạn không dịch được

4 Bảng Morse chuẩn

Alphanumeric characters.

.- A --. G -- M ... S -.—Y 5

| | | | | | |
|--------|--------|--------|--------|---------|---------|
| -. . B | H | -. N | - T | --.. Z | -.... 6 |
| -. . C | .. I | --- O | ..- U | .---- 1 | --... 7 |
| -. . D | .--- J | -. . P | ...- V | .---- 2 | ---.. 8 |
| . E | -. . K | --- Q | -. . W | ...—3 | ----. 9 |
| -. . F | -. . L | -. . R | -. . X |- 4 | ----- 0 |

Các dấu

Point (.) .-..- (AAA)

Comma (,) --..-- (MIM)

Question mark (?) ..--- (IMI)

Colon (:) ---... (OS)

Hyphen (-) -....- (BA)

Error

5 Ví dụ

Ví dụ là file dữ liệu myinput.dat và đầu ra là file conver.txt Bạn sẽ viết

```
:> morse myinput.dat conver.txt
```

Đây là một đoạn dữ liệu lấy từ khi mã Morse mới bắt đầu năm 1844 như sau:

What hath God wrought

với mã tương đương là

```
.-- .... .- - / .... .- - .... / --. --- -. / .-- .- --- .. ---. .... -
```

và sau khi chuyển ngược lại thành dòng lệnh

```
what hath god wrought
```

Tương tự với dòng lệnh sau cho xử lý lỗi:

```
-. . .... .- -/--- -. /.- / -- .- - ..--- .-.-  
-. . .- .-.-.-.- --- / .. .- / .-/ .- .- --- .-.-.-
```

sẽ tạo ra file text :

```
c#at on a mat*.
```

```
bu*g in a rug.
```

và hiện thị thông báo lỗi

```
Error: Invalid Morse code .-.-.. on line 1.
```

```
Error: Invalid Morse code .-.-.-.- on line 2.
```

6 Thiết kế

Phương pháp thiết kế truyền thống (top – down). Bạn sẽ phải vẽ sơ đồ thiết kế hình cây và đưa ra giải thích về sự chuyển đổi thông tin giữa các phần của chương trình.

7 Code

Lưu ý, để dễ dàng cho việc chấm, mỗi hàm cần có lời giải thích ngắn gọn, đầu vào là gì, đầu ra là gì. Bạn nên đưa ra giải thích trong hàm khi có thể. Lưu ý việc sử dụng tên hàm, tên biến cho dễ hiểu và thống nhất. Các hàm phải cách nhau bằng hàng trống, tránh sử dụng các giá trị hằng. Trong chương trình có nhiều file, các file header luôn giải thích rõ ràng để giúp người đọc hiểu rõ công dụng của các file này.

8 Gợi ý

Sử dụng các hàm sau khi làm việc với file
`fgetc()`, `fgets()`, `ungetc()`, `fputc()`, `fprintf()`, and `rewind()`.

Để kiểm tra xem file đầu ra đã có chưa, bạn nên thử mở file đó, nếu hàm trả về con trỏ `!=NULL`, thì file đó đã xuất hiện.

Bạn có thể giải bài toán này bằng nhiều cách, nhưng tốt nhất là lưu dữ liệu thành các mảng cấu trúc và duyệt qua cấu trúc này khi tìm kiếm. Việc này sẽ làm cho chương trình của bạn dễ dàng mở rộng hơn.

Chương trình nên bắt đầu bằng việc duyệt qua những file nguồn đơn giản, trước khi bạn viết ra các phiên bản mở rộng hơn, trong đó xét đến những vấn đề xử lý lỗi của chương trình.

Yêu cầu bài tập dài số 4

1. Thời hạn nộp bài tập dài: 1 tuần
2. Địa điểm: Nộp Online
3. Phần mềm nộp được nộp bao gồm
 - a. Báo cáo về bài tập dài
 - b. Chương trình phần mềm chạy bài tập gồm file mã nguồn và file chạy chương trình (.exe) viết trên Win32.
 - c. Mỗi thư mục ghi rõ: Họ và tên sinh viên, lớp và số hiệu SV
4. Bài tập dài sẽ được đánh giá trên hai điểm chính:
 - a. Chương trình chạy có đáp ứng được các yêu cầu bài tập đặt ra không?
 - b. Mã nguồn của chương trình có được viết tốt không?
5. Nội dung của báo cáo cho bài tập dài số 1 bao gồm những phần sau:
 - a. Những yêu cầu nào mà phần mềm của anh (chị) sẽ giải quyết, bao gồm các đầu vào của chương trình là thế nào ? Chương trình sẽ không hoạt động khi cho vào những dữ liệu kiểu thế nào ? (Lưu ý không tìm cách thêm chức năng cho phần mềm).
 - b. Cấu trúc cơ bản của phần mềm viết ra theo sơ đồ hình cây (top down), đồng thời giải thích rõ các cấu trúc dữ liệu chính trong phần mềm .
 - c. Những điểm theo anh (chị) là quan trọng quyết định đến hiệu quả của phần mềm này
 - d. Bản báo cáo sẽ dài tối đa là 5 trang (ngoại trừ hình vẽ) và không bao gồm mã nguồn của chương trình.
6. Lưu ý
 - a. Đây là bài tập dài dành cho từng cá nhân và mọi người phải tự làm toàn bộ các báo cáo và phần mềm của riêng mình. Mọi sao chép (dù chỉ một phần) trong báo cáo hay trong phần mềm đều bị coi là phạm quy và bài viết sẽ được cho điểm 0.