

# VIQR\_UTF8 Converter

## Tổng quan chương trình

VIQR\_UTF8 Converter là Console Application để đọc văn bản tiếng Việt được mã hóa bằng VIQR từ một tệp đầu vào và xuất ra dưới dạng UTF8 tương ứng vào một tệp đầu ra, và ngược lại.

## Student Information

- Họ và tên: Cao Hoài Việt
- MSSV: 22850034
- Email: [22850034@student.hcmus.edu.vn](mailto:22850034@student.hcmus.edu.vn)
- Github Repository: [https://github.com/vietch2612/22850034-hcmus-assignment/tree/main/data\\_organization/ex4](https://github.com/vietch2612/22850034-hcmus-assignment/tree/main/data_organization/ex4)

Tự đánh giá mức độ hoàn thành: 100%

Task	Status
Code	✓
Build on macOS & Windows	✓
Unit test	✓
Documentation	✓

## Công nghệ sử dụng

- C++11
- CMake <https://cmake.org/>
- Google Test <https://github.com/google/googletest>

## Yêu cầu

Với Windows:

- Cài đặt CMake: <https://cmake.org/download/>

## Với Linux (Ubuntu):

```
sudo apt-get install cmake
```

## Với macOS:

```
brew install cmake
```

## Biên dịch

---

1. Giải nén các tệp tin.
2. Mở terminal và cd vào thư mục chứa project.
3. Biên dịch ứng dụng bằng CMake:

Với macOS và Linux

```
cd VIQR_UTF8
mkdir build
cd build
cmake ..
cmake --build . --config Release
```

Với Windows (PowerShell)

```
cd VIQR_UTF8
mkdir build
cd build
cmake .. -G"Visual Studio 17 2022"
cmake --build . --config Release
```

## Chạy ứng dụng

---

Với macOS và Linux

```
# Run inside the build folder
# Init the in.txt file
echo "Vie^t Nam dda^t nu*o*'c me^'n ye^u\." > in.txt
# Run the application
./VIQR_UTF8 1 in.txt out.txt
```

```
# View the result
cat out.txt
```

## Với Windows (PowerShell)

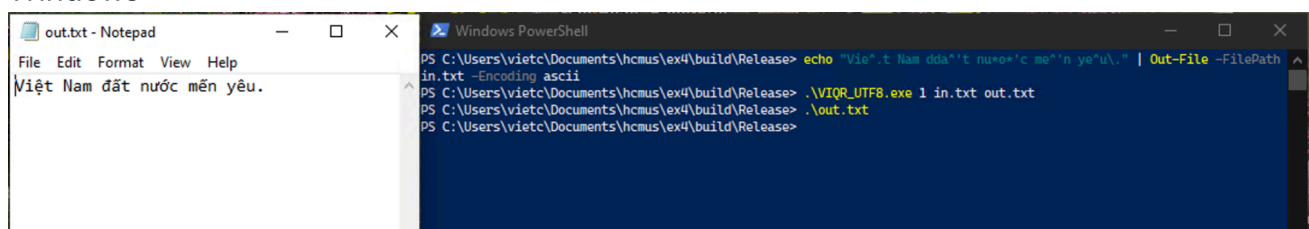
```
# Open the Release folder inside the build folder
cd Release
# Init the in.txt file with ASCII encoding
echo "Vie^.t Nam dda^'t nu*o*'c me^'n ye^u\." | Out-File -FilePath in.txt -
Encoding ascii
.\VIQR_UTF8.exe 1 in.txt out.txt
# View the result
.\out.txt
# Convert back to VIQR
./VIQR_UTF8.exe 2 out.txt in2.txt
# View the result
cat in2.txt
```

## Kết quả chạy chương trình

### macOS

```
● build >>> echo "Vie^.t Nam dda^'t nu*o*'c me^'n ye^u\." > in.txt
● build >>> ./VIQR_UTF8 1 in.txt out.txt
● build >>> cat out.txt
Việt Nam đất nước mến yêu.
● build >>> cat in.txt
Vie^.t Nam dda^'t nu*o*'c me^'n ye^u\.
● build >>> ./VIQR_UTF8 2 out.txt in2.txt
● build >>> cat in2.txt
Việt Nam đất nước mến yêu.
○ build >>> █
```

### Windows



## Cấu trúc chương trình

VIQR\_UTF8/

```
|
| — build/ # Chứa các tệp tin biên dịch
| — Release/ # Chứa tệp tin biên dịch cho phiên bản Release (.exe với
Windows)
```

```
| |— (generated files) # Các tệp tin được tạo ra trong quá trình biên dịch
|— CMakeLists.txt # Tệp tin CMake cho project, chứa các hướng dẫn cho
trình biên dịch CMake
|— FileProcessor.h # Tệp header cho lớp FileProcessor, chứa khai báo các
phương thức
|— FileProcessor.cpp # Triển khai các phương thức của lớp FileProcessor
|— FileProcessor_test.cpp # Các test cho lớp FileProcessor
|— main.cpp # Hàm main của chương trình
|— README.md # Thông tin và tài liệu dự án bằng tiếng Anh
|— README_Vietnamese.md # Thông tin và tài liệu dự án bằng tiếng Việt
|— VIQR_UTF8.h # Tệp header cho lớp VIQR_UTF8, chứa khai báo các phương
thức
|— VIQR_UTF8.cpp # Triển khai các phương thức của lớp VIQR_UTF8
|— VIQR_UTF8_test.cpp # Các test cho chức năng chuyển đổi VIQR sang UTF-8
|— screenshots/ # Chứa các ảnh chụp màn hình dùng cho README
| |— *.png # Các ảnh chụp màn hình
```

1. Lớp `FileProcessor` chứa các phương thức để đọc và ghi dữ liệu từ tệp tin (txt). Nó bao gồm:
  - `readFile` Đọc dữ liệu
  - `writeFile` Ghi dữ liệu
2. Lớp `VIQR_UTF8` chứa các phương thức để chuyển đổi văn bản giữa VIQR và UTF-8. Nó bao gồm:
  - Vectors `viqrChars` và `utf8Chars` chứa các ký tự VIQR và UTF-8 tương ứng.
  - `viqrToUtf8` Chuyển đổi từ VIQR sang UTF-8
  - `utf8ToViqr` Chuyển đổi từ UTF-8 sang VIQR

## Cách thức hoạt động

1. `ProcessFile` đọc dữ liệu từ tệp tin đầu vào (txt) và gọi tới `viqrToUtf8` hoặc `utf8ToViqr` trong lớp `VIQR_UTF8` để chuyển đổi dữ liệu dựa vào giá trị của `mode`.
2. Hàm `viqrToUtf8` chuyển đổi văn bản VIQR sang UTF-8.
  - The first loop reads the input text character by character.
  - Vòng lặp đầu tiên đọc dữ liệu đầu vào, từng ký tự một.
  - Vòng lặp thứ hai kiểm tra xem ký tự đó có phải là ký tự VIQR hay không bằng cách so sánh từng ký tự VIQR trong vector `viqrChars` với ký tự hiện tại tính từ `i` tới `i+viqrChars[j].length()`.
  - Nếu ký tự đó là ký tự VIQR, hàm sẽ thay thế nó bằng ký tự UTF-8 tương ứng.
  - Nếu ký tự đó không phải là ký tự VIQR, hàm sẽ thêm nó vào file kết quả.
3. Hàm `utf8ToViqr` chuyển đổi văn bản UTF-8 sang VIQR.

- Phương pháp tương tự như hàm `viqrToUtf8` nhưng ngược lại. Và hàm này chỉ cần kiểm tra từng kí tự một, không cần phải kiểm tra theo độ dài của kí tự VIQR.

## Unit tests

---

Chương trình sử dụng Google Test để kiểm thử đơn vị (Unit test). Các test được định nghĩa trong các tệp `FileProcessor_test.cpp` và `VIQR_UTF8_test.cpp`. Kiểm thử đơn vị giúp đảm bảo rằng các hàm hoạt động đúng như mong đợi mỗi khi có thay đổi trong code.

Để chạy test, biên dịch project và chạy các lệnh sau:

```
cd build
cmake --build . && ctest
```

Danh sách các test case được định nghĩa để bao phủ tất cả\* các trường hợp có thể:

(\*) *Tất cả các trường hợp em đã nghĩ tới.*

- `VIQRToUtf8Conversion.BaseCharacters`
- `VIQRToUtf8Conversion.SpecialCharacters`
- `VIQRToUtf8Conversion.CharactersWithDiacritics`
- `VIQRToUtf8Conversion.FullParagraph`
- `VIQRToUtf8Conversion.EscapeCharacters`
- `VIQRToUtf8Conversion.UpperCaseCharacters`
- `Utf8ToViqrConversion.BaseCharacters`
- `Utf8ToViqrConversion.SpecialCharacters`
- `Utf8ToViqrConversion.CharactersWithDiacritics`
- `Utf8ToViqrConversion.FullParagraph`
- `Utf8ToViqrConversion.EscapeCharacters`
- `ProcessFileTest.ViqrToUtf8Conversion`
- `ProcessFileTest.Utf8ToViqrConversion`
- `ProcessFileTest.EscapeCharacters`

## Kết quả test:

```

● data_organization >>> cd build
● build >>> cmake --build . && ctest
[ 6%] Building CXX object CMakeFiles/VIQR_UTF8_Lib.dir/VIQR_UTF8.cpp.o
[ 12%] Building CXX object CMakeFiles/VIQR_UTF8_Lib.dir/FileProcessor.cpp.o
[ 18%] Linking CXX static library libVIQR_UTF8_Lib.a
[ 18%] Built target VIQR_UTF8_Lib
[ 25%] Building CXX object CMakeFiles/VIQR_UTF8.dir/main.cpp.o
[ 31%] Linking CXX executable VIQR_UTF8
[ 31%] Built target VIQR_UTF8
[ 37%] Building CXX object _deps/googletest-build/googletest/CMakeFiles/gtest.dir/src/gtest-all.cc.o
[ 43%] Linking CXX static library ../../lib/libgtest.a
[ 43%] Built target gtest
[ 50%] Building CXX object _deps/googletest-build/googletest/CMakeFiles/gtest_main.dir/src/gtest_main.cc.o
[ 56%] Linking CXX static library ../../lib/libgtest_main.a
[ 56%] Built target gtest_main
[ 62%] Building CXX object CMakeFiles/VIQR_UTF8_Test.dir/VIQR_UTF8_test.cpp.o
[ 68%] Building CXX object CMakeFiles/VIQR_UTF8_Test.dir/FileProcessor_test.cpp.o
[ 75%] Linking CXX executable VIQR_UTF8_Test
[ 75%] Built target VIQR_UTF8_Test
[ 81%] Building CXX object _deps/googletest-build/googmock/CMakeFiles/gmock.dir/src/gmock-all.cc.o
[ 87%] Linking CXX static library ../../lib/libgmock.a
[ 87%] Built target gmock
[ 93%] Building CXX object _deps/googletest-build/googmock/CMakeFiles/gmock_main.dir/src/gmock_main.cc.o
[100%] Linking CXX static library ../../lib/libgmock_main.a
[100%] Built target gmock_main
Test project /Users/vietcao/repos/hcmus-assignment/data_organization/build
  Start 1: VIQRToUtf8Conversion.BaseCharacters
1/14 Test #1: VIQRToUtf8Conversion.BaseCharacters ..... Passed    0.00 sec
  Start 2: VIQRToUtf8Conversion.SpecialCharacters
2/14 Test #2: VIQRToUtf8Conversion.SpecialCharacters ..... Passed    0.00 sec
  Start 3: VIQRToUtf8Conversion.CharactersWithDiacritics
3/14 Test #3: VIQRToUtf8Conversion.CharactersWithDiacritics ... Passed    0.00 sec
  Start 4: VIQRToUtf8Conversion.FullParagraph
4/14 Test #4: VIQRToUtf8Conversion.FullParagraph ..... Passed    0.00 sec
  Start 5: VIQRToUtf8Conversion.EscapeCharacters
5/14 Test #5: VIQRToUtf8Conversion.EscapeCharacters ..... Passed    0.00 sec
  Start 6: VIQRToUtf8Conversion.UpperCaseCharacters
6/14 Test #6: VIQRToUtf8Conversion.UpperCaseCharacters ..... Passed    0.00 sec
  Start 7: Utf8ToViqrConversion.BaseCharacters
7/14 Test #7: Utf8ToViqrConversion.BaseCharacters ..... Passed    0.00 sec
  Start 8: Utf8ToViqrConversion.SpecialCharacters
8/14 Test #8: Utf8ToViqrConversion.SpecialCharacters ..... Passed    0.00 sec
  Start 9: Utf8ToViqrConversion.CharactersWithDiacritics
9/14 Test #9: Utf8ToViqrConversion.CharactersWithDiacritics ... Passed    0.00 sec
  Start 10: Utf8ToViqrConversion.FullParagraph
10/14 Test #10: Utf8ToViqrConversion.FullParagraph ..... Passed    0.00 sec
  Start 11: Utf8ToViqrConversion.EscapeCharacters
11/14 Test #11: Utf8ToViqrConversion.EscapeCharacters ..... Passed    0.00 sec
  Start 12: ProcessFileTest.ViqrToUtf8Conversion
12/14 Test #12: ProcessFileTest.ViqrToUtf8Conversion ..... Passed    0.00 sec
  Start 13: ProcessFileTest.Utf8ToViqrConversion
13/14 Test #13: ProcessFileTest.Utf8ToViqrConversion ..... Passed    0.00 sec
  Start 14: ProcessFileTest.EscapeCharacters
14/14 Test #14: ProcessFileTest.EscapeCharacters ..... Passed    0.00 sec

100% tests passed, 0 tests failed out of 14

Total Test time (real) = 0.05 sec
○ build >>> █

```