CIS–3152 Lab
Character Conversions

Peter C. Chapin*
Vermont Technical College

Last Revised: March 2, 2013

Introduction

In this lab you will implement a short Java program that does character encoding conversions. You will then use this program to encode the same data several different ways and observe the differences between the encoded forms. Specifically, you will experiment with UTF-8, UTF-16, and ISO-8859-1.

1 The Program

In the samples repository for this course you will find a skeletal Java conversion program named `CharConverter`. The Java libraries provide classes that know how to decode into Java's internal character representation (essentially UTF-16) external files that have been encoded in various ways. Similarly there are library classes that can encode Java characters into various forms for saving to disk.

The `CharConverter` program is to use these libraries to convert a file in one encoding into a file in another. For example

```
$ java -jar CharConversions.jar \
    input.txt:UTF-8 output.txt:UTF-16BE
```

converts the file `input.txt` using UTF-8 into the file `output.txt` using UTF-16BE. Notice how the name of the encoding follows the name of the file, delimited by a colon character. This is the syntax the `CharConverter` program expects on its command line.

You should begin by completing the skeleton. Follow the comments in the source file and be sure to browse the Java SDK documentation to learn about the various classes and methods used.

2 Experiments

Also in the samples repository you will find a file `sample.utf8` that contains some text in UTF-8. Use the program above to convert this file into UTF-16BE, UTF-16LE, and ISO-8859. Take note of the size of each converted result and also the first few bytes of each file (at least up through the first word). Are the results what you would expect?

3 Report

For this lab you only need to submit a commented copy of your final program along with some informal notes reporting on the results of your experiments.

*PChapin@vtc.vsc.edu