

COMP202-08B Computer Communications

Lecture 8a A note on dealing with binary data



7 August 2008

Data

- Binary data consists of bytes
 - Values 0-255
- A text file consists of a subset of these bytes
 - Printable with an ascii symbol
 - A-Z, a-z, 0-9, punctuation, etc
 - Carriage return, newline, tabs, spaces.

© THE UNIVERSITY OF WAIKATO - TE WHARE WAIKATO O HOKIANGA

7 August 2008

2

End of line sequence

- The sequence of bytes used to get to the next line depends on the operating system
 - Windows: `\r\n` CR LF
 - UNIX: `\n` LF
 - MacOS < 10: `\r` CR
- CR byte value 13
- LF byte value 10

© THE UNIVERSITY OF WAIKATO - TE WHARE WAIKATO O HOKIANGA

7 August 2008

3

readLine

- Java's `readLine` function will handle any combination of these symbols when reading lines of Strings from an `InputStream`
 - CR `\r` 13
 - LF `\n` 10
 - CR LF `\r\n` 13 10
- Java's `println` function will not necessarily emit the same sequence of bytes when you `println`

© THE UNIVERSITY OF WAIKATO - TE WHARE WAIKATO O HOKIANGA

7 August 2008

4

Lost in translation

```
    H   E   L   L   O   \r   \n
                        ↓ readLine()
String val  H   E   L   L   O
                        ↓ println()
          H   E   L   L   O   \n
```

© THE UNIVERSITY OF WAIKATO - TE WHARE WAIKATO O HOKIANGA

7 August 2008

5

Binary Files

- Binary files will contain bytes, the exact interpretation of which depends on the type of file
 - Image file
 - Program file
- There will be bytes with values 10 and 13 sprinkled amongst the file
 - They do not mean `\n` or `\r` in this situation

© THE UNIVERSITY OF WAIKATO - TE WHARE WAIKATO O HOKIANGA

7 August 2008

6

Binary files and readLine

- Binary files will almost certainly lose some CR or LF characters when using readLine / println
- Need to read binary files into a byte array
 - `byte[] buf = new byte[1024];`
 - `FileInputStream fis = new FileInputStream(/* something */);`
 - `int rc = fis.read(buf);`
- Write bytes to OutputStream

HTTP

- HTTP headers are printable ascii, but with a specific end-of-line sequence
 - CR LF
- The HTTP response will include an ascii header, followed by binary data.
 - You probably do not want to write the ascii header using println, though.