Unicode Standard

Outline

1. Unicode Standard
2. ANSI escape code
3. Escape Sequences
4. Project
5. Interview Questions
6. Summary

Unicode Standard

Before Unicode, there were many language standards

- ASCII
- ISO 8859-1
- KOI-8
- GB18030 and BIG-5

Problem

- A particular code value corresponds to different letters in the various language standards.
- The encodings for languages with large character sets have variable length.Some common characters are encoded as single bytes, other require two or more byte.

Unicode Standard

- Unicode, international character-encoding system. The Unicode Standard includes letters, digits, diacritics, punctuation marks, technical symbols, emoji and other symbols, using a uniform encoding scheme. The standard is maintained by the Unicode Consortium. The first version of Unicode was introduced in 1991; the most recent version contains more than 100,000 characters.
- The Unicode standard uses hexadecimal to express a character. The Unicode standard was initially designed using 16 bits to encode characters because the primary machines were 16-bit PCs.
- When the specification for the Java language was created, the Unicode standard was accepted and the char primitive was defined as a 16-bit data type, with characters in the hexadecimal range from 0x0000 to 0xFFFF.
- **Surrogates**
- Unicode home
- Unicode table

ANSI escape code

ANSI escape sequences are a standard for in-band signaling to control cursor location, color, font styling, and other options on video text terminals and terminal emulators. Certain sequences of bytes, most starting with an ASCII escape character and a bracket character, are embedded into text. The terminal interprets these sequences as commands, rather than text to display verbatim.

```
ANSI_RESET  = "\u001B[0m"
ANSI_BLACK  = "\u001B[30m"
ANSI_RED    = "\u001B[31m"
ANSI_GREEN  = "\u001B[32m"
ANSI_YELLOW = "\u001B[33m"
ANSI_BLUE   = "\u001B[34m"
ANSI_PURPLE = "\u001B[35m"
ANSI_CYAN   = "\u001B[36m"
ANSI_WHITE  = "\u001B[37m"
```

Github Link

Escape Sequences

A character preceded by a backslash (\) is an *escape sequence* and has special meaning to the compiler.

\t	Insert a tab in the text at this point.
\n	Insert a newline in the text at this point.

Oracle Documentation

Interview Questions

Hard work beats talent when talent doesn’t work hard.