6.6 Surrogates Area

The Surrogates Area of the Unicode Standard provides 2,048 character codes that are used in the surrogate extension method (see Figure 6-31).

Figure 6-31. Surrogates
Surrogates Area: U+D800—U+DFFF

The Surrogates Area consists of 1,024 low-half surrogate code values and 1,024 high-half surrogate code values, which are interpreted in pairs to access over a million code points. Surrogate-pairs are designed to allow representation of rare characters in future extensions of the Unicode Standard. There are no such characters currently assigned in this version of this standard. (For the formal definition of a surrogate-pair and the role of surrogate-pairs in the Unicode Conformance Clause, see Section 3.7, Surrogates.)

The use of surrogate-pairs in the Unicode Standard is formally equivalent to the Universal Transformation Format-16 (UTF-16) defined in ISO 10646. (For a complete statement of the UTF-16 extension mechanism, see Appendix C, Relationship to ISO/IEC 10646.)

**High Surrogate.** The high surrogate code values are assigned to the range U+D800 → U+DBFF. The high surrogate code value is always the first element of a surrogate-pair.

**Low Surrogate.** The low surrogate code values are assigned to the range U+DC00 → U+DFFF. The low surrogate code value is always the second element of a surrogate-pair.

**Private Use High Surrogates.** The high surrogate code values from U+DB80 → U+DBFF are private use high surrogate code values (a total of 128 code values). Characters represented by means of a surrogate-pair, where the high surrogate code value is a private use high surrogate, are private use characters. This mechanism allows for a total of 131,072 (i.e., 2^21) private use characters representable by means of surrogate-pairs. (For more information on private use characters, see the discussion of the Private Use Area.)

**Encoding Structure.** The Surrogate Area is divided into the following ranges:

```
U+D800  →  U+DB7F  High surrogates
U+DB80  →  U+DBFF  Private Use high surrogates
U+DC00  →  U+DFFF  Low surrogates
```

⇒ There are no charts for this area, as no surrogate-pairs have assigned characters associated with them.