Sorting

String sorting is used in a number of places in the Macintosh Operating System (for example, in a standard file dialog box) and in applications (for instance, spreadsheets). When performing such sorting, it is important to order strings in the manner expected by the user-that is, according to the rules of the language and region for which the system is localized. The **International Utilities Package** provides several routines that compare two strings and indicate whether the first should be sorted before, after, or at the same place as the second string. For details, see **Using the International Utilities Package Routines**.

Sorting or comparing strings can be an extremely intricate operation. Subtle issues like expansion, contraction, ignorable characters, and exceptional words may be taken into account. Sorting cannot be done properly by a simple table look-up, even for such straightforward cases as English. Sorting depends not just on the script, but on the individual language. While broad similarities in sorting exist between languages that share the same script, definite variations between languages must be taken into account.

The **Script Manager**, the **International Utilities Package**, and international resource 'itl2' have long provided for many sorting issues, including primary or secondary order, expansion, contraction, and ignorable characters. With system software version 7.0, several new sorting capabilities provide support for systems with multiple installed scripts and language capabilities.

- You can sort strings in different scripts and languages.
- A new international resource, 'itlm', indicates the preferred sorting order for scripts, languages, and region codes, and indicates how to map region codes to languages and language codes to scripts. See **The 'itlm' Resource** for details.
- You can explicitly specify the handle of the resource to be used for sorting. This is helpful for multilingual systems. See the routines **IUCompPString**, **IUMagPString**, **IUEqualPString**, and **IUMagIDPString** for details.
- 'itl2' and 'itl4' resource handles for all active scripts are cached by the **Script Manager**. You can call a routine to clear the cache so application-supplied resources can be used. See the section entitled, **Accessing the International Resources** for details.

Primary or Secondary Order

Sorting order is determined by a ranking of the entire standard Roman character set. This ranking can be thought of as a two-dimensional table. Each row is a class of characters-for example, all of the forms of uppercase and lowercase *A* with and without various diacritical marks. The characters are ordered within the row, but that ordering is secondary to the primary ordering of the rows themselves. For example, all of the forms of *A* precede all of the forms of *B,* as follows:

A < Å <a < å

B < b

Primary sorting characteristics denote a strong ranking; if any primary differences are present, all secondary differences are ignored. For instance, only primary sorting is needed to determine that *abc* precedes *bc*. Secondary sorting characteristics indicate that if certain differences are present, a second pass is made that introduces a weak ordering. Here's an example:

abc < åbc

Expansion

A single character may be sorted as if it were a sequence of characters. First, the single character is expanded; then the primary sorting occurs based on this expansion. In the secondary sorting, the characters are recombined. For instance, *ä* in German may be sorted as if it were the two characters *ae,* as in this example:

bäk < baek < bäks

Contraction

A sequence of characters may be sorted as a single character. For instance, *ch* in Spanish may be sorted as if it were one character that sorts after *c,* as in this example:

czar < char< dar

Ignorable Characters

Certain characters should be ignored unless the strings are otherwise equal. In other words, they have no effect on primary sorting, but they do influence secondary sorting. Examples of ignorable characters in English are hyphens, apostrophes, and spaces. Here is an example of how a hyphen influences secondary sorting:

blackbird < black-bird < blackbirds

Exceptional Words

Sometimes the sorting order changes drastically for special cases. For instance, when words are understood to be abbreviations, the strings are sorted as if they were spelled out.

McDonald < Mary McDonald is treated as MacDonald
Thus, MacDonald < Mary

St. James < Smith St. is an abbreviation for Saint h
Saint James < Smith

Easy Step < Easy St. St. is an abbreviation for Street n
Easy Step < Easy Street

Such cases require a direct dictionary look-up and are not handled by the Macintosh Script Management System. Note that abbreviations are

context-dependent; for example, *St*. may denote *Saint* or *Street,* depending on the meaning of the adjacent text.