

What is a Charset and what is it made of?

Each piece of software works with a character set. A character set is a set of symbols, including letters, digits, spaces, and other symbols. A character set is a collection of encodings and symbols. A character set's character comparisons can be done using a set of rules called a collation.

For example, let's say we take the letters from a to z from the alphabet. Each letter gets a number in order; a = 0, b = 1, c = 2, etc... The letter **a is the Symbol** and **the number 0 is the encoding** for the letter A. The combination of letters and encodings form a character set.

Character Set

Firebase and InterBase both support more than 20 different character sets for use around the world. When building a database, if you don't specify a character, the IBExpert program will automatically use the default character set it automatically assumes. The chosen character set is very important when importing and exporting data with different character sets. This needs to be taken into consideration when applications are developed with multiple language versions.

- The **ASCII character set** (American Standard Code for Information Interchange): If no character set is defined, Firebird or InterBase chooses the character set NONE, and values are stored exactly as typed. A column defined with NONE can accept data from any character set, but the same data cannot be loaded into a column defined with a different character set. All characters are converted into their ASCII equivalents from the character set they were input under when the ASCII character set is specified.
- The **UTF-8 character set** (UCS Transformation Format 8): A Unicode-based encoding such as UTF-8 covers almost all of the characters and symbols in the world. Its use also eliminates the need for server-side logic to individually determine the character encoding for each page served or each incoming form submission
- If you use **ANSI character set** (American National Standards Institute) to work with your databases there are no changes to previous versions of IBExpert, except it is now possible to enter characters that are not presented in your default system locale. Such characters will be replaced by converting from Unicode to ANSI representation.

It is crucial to specify the appropriate default character set for your application and requirements.

Change Character Set

Generally this default character set cannot be altered at a later date (only using the command line tool IBEScript). Alternate character sets can however be defined for individual domains and tables, which override the default character set.

From:
<http://ibexpert.com/docu/> - **IBExpert**

Permanent link:
<http://ibexpert.com/docu/doku.php?id=02-ibexpert:02-01-getting-started:registering-a-database:character-sets>

Last update: **2023/05/25 04:15**

