

# Overview of the main character sets in Firebird

By Stefan Heymann

Character sets are an issue every programmer has to deal with one day. This is an overview of the most important character sets.

Name	Bytes per Character	Description	Range	IANA/MIME Code
7-bit ASCII	1	The mother of all character sets. Contains 32 invisible control characters, the Latin letters A-Z, a-z, the Arabic digits 0-9 and a bunch of punctual characters. Code Range 0..127.	0..127	US-ASCII

## Unicode-based Character Sets

Unicode, ISO 10646	N.A.	A universal code for all characters anyone can think of. Defines characters, assigns them a scalar value, but does not define how characters are rendered graphically or in memory.	U+0000..U+100000	N.A.
UTF-8	1..6	A Unicode transformation format which uses 1-Byte characters for all 7-bit US-ASCII characters and sequences of up to 6 bytes for all other Unicode characters.	All Unicode characters	UTF-8
UCS-2	2	A unicode transformation format which uses 2 Bytes (16 Bits) for every character. This character set is not able to render all Unicode scalars and is therefore obsolete. However, it is still used by a lot of systems (Java, NT)	U+0000..U+FFFF	ISO-10646-UCS-2
UTF-16	2	A unicode transformation format which uses 2 Bytes (16 Bits) for every character. Using the concept of "Surrogate Pairs", this format is able to render all Unicode characters.	All Unicode characters	UTF-16
UCS-4, UTF-32	4	Two unicode transformation formats which use 4 Bytes (32 Bits) for every character. UCS-4 and UTF-32 are the only character sets, which are able to render all Unicode characters in equally long words. UCS-4 and UTF-32 are technically identical.	All Unicode characters	ISO-10646-UCS-4, UTF-32

## Single-byte Character Sets

ISO 8859-x	1	An extension of US-ASCII using the eighth bit.	0..127, 160..255	ISO-8859-x
Windows 125x	1	Equal to ISO 8859-x, plus additional characters in the 128..159 range.	0..255	Windows-125x

## ISO 8859-x Character Sets

Name		Covered Languages	MS Windows counterpart
ISO 8859-1	Latin-1	Western and West European languages (English, German, French, Spanish, Portuguese, etc.). As these languages are used in large parts of the world (Europe, Americas, Australia, Africa), these are the most widely used character sets.	Windows-1252
		Windows 1252 and ISO 8895-1 are equal in the 160..255 range.	Windows-1252
ISO 8859-2	Latin-2	Central and East European languages (Czech, Polish, etc.).	Windows-1250
ISO 8859-3	Latin-3	South European, Maltese, Esperanto	
ISO 8859-4	Latin-4	North European	
ISO 8859-5	Cyrillic	Russian, Ukrainian	Windows-1251
ISO 8859-6	Arabic	Arabic	Windows-1256
ISO 8859-7	Greek	Modern Greek	Windows-1253
ISO 8859-8	Hebrew	Hebrew	Windows-1255
ISO 8859-9	Latin-5	Turkish	Windows-1254
ISO 8859-10	Latin-6	Nordic (Sami, Inuit, Icelandic)	
ISO 8859-11	Thai	Thai	Windows-874
ISO 8859-13	Latin-7	Baltic	Windows-1257
ISO 8859-14	Latin-8	Celtic	
ISO 8859-15	Latin-9	Similar to ISO 8859-1, adds Euro sign (€) and a few other characters.	
ISO 8859-16	Latin-10	South Eastern European languages (Albanian, Croatian, Hungarian, Italian, Polish, Romanian, Slovenian, but also Finnish, French, German and Irish Gaelic).	

## MS Windows Character Sets

Number	Name
1250	Latin 2
1251	Cyrillic
1252	Latin 1
1253	Greek
1254	Latin 5
1255	Hebrew
1256	Arabic
1257	Baltic
1258	Viet Nam
874	Thai

Last updated 2010-02-23

For a complete list of the character sets available for your database version, connect to your database, and take a look at the list of character sets in the RDB\$CHARACTER\_SETS system table:

Table : [RDB\$CHARACTER\_SETS] : Employee\_2\_5 (localhost:C:\FB3025\examples\empbuild\EMPLOYEE.FDB)

Fields Constraints Indices Dependencies Triggers Data Master/Detail View Description DDL Grants Comparison To-do

Record: 45 52 records fetched

RDB\$CHARACTER_SET_N...	RDB\$FO...	RDB\$N...	RDB\$DEFAULT_COLLATE_NAME	RDB\$CHARACTER_SET_ID	RDB\$BYTES_PER_CHARACTER	R...	RDE
ISO8859_3	<null>	<null>	ISO8859_3	23	1	1	<nu
ISO8859_4	<null>	<null>	ISO8859_4	34	1	1	<nu
ISO8859_5	<null>	<null>	ISO8859_5	35	1	1	<nu
ISO8859_6	<null>	<null>	ISO8859_6	36	1	1	<nu
ISO8859_7	<null>	<null>	ISO8859_7	37	1	1	<nu
ISO8859_8	<null>	<null>	ISO8859_8	38	1	1	<nu
ISO8859_9	<null>	<null>	ISO8859_9	39	1	1	<nu
KOIBR	<null>	<null>	KOIBR	63	1	1	<nu
KOIBU	<null>	<null>	KOIBU	64	1	1	<nu
KSC_5601	<null>	<null>	KSC_5601	44	2	1	<nu
NEXT	<null>	<null>	NEXT	19	1	1	<nu
NONE	<null>	<null>	NONE	0	1	1	<nu
OCTETS	<null>	<null>	OCTETS	1	1	1	<nu
SJIS_0208	<null>	<null>	SJIS_0208	5	2	1	<nu
TIS620	<null>	<null>	TIS620	66	1	1	<nu
UNICODE_FSS	<null>	<null>	UNICODE_FSS	3	3	1	<nu
UTF8	<null>	<null>	UTF8	4	4	1	<nu
WIN1250	<null>	<null>	WIN1250	51	1	1	<nu
WIN1251	<null>	<null>	WIN1251	52	1	1	<nu
WIN1252	<null>	<null>	WIN1252	53	1	1	<nu
WIN1253	<null>	<null>	WIN1253	54	1	1	<nu

Grid View Form View Print Data

See also:

- [Default character set](#)
- [Character set](#)
- [Character sets and Unicode in Firebird](#)
- [Convert your Firebird applications to Unicode](#)
- [InterBase® character sets and collation orders](#)
- [New character sets in Firebird 2.5](#)
- [New character sets in Firebird 2.1](#)
- [New character sets in Firebird 2.0](#)

[back to top of page](#)

Any comments? Send an email to [register@ibexpert.biz](mailto:register@ibexpert.biz)

COPYRIGHT © 2002-2022 HK-Software, IBExpert Ltd. All rights reserved.

All IBExpert brand and product names are trademarks or registered trademarks of IBExpert Ltd in Malta and other countries. InterBase, Delphi, CodeGear, C++Builder, Delphi/400, Delphi for PHP and JBuilder are trademarks or registered trademarks of Embarcadero Technologies Inc. in the United States and other countries.

Firebird is a registered trademark of the FirebirdSQL Foundation. Turbo Pascal is a registered trademark of Borland International, Inc. Sun, Java, JavaScript and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark in the United States and other countries, exclusively licensed through "The Open Group". Oracle is a registered trademark of Oracle Corporation in the United States and

other countries. All Microsoft brand and product names are trademarks or registered trademarks of Microsoft Corporation in the United States and other countries.

AS/400, DB2, IBM, Informix and iSeries are trademarks or registered trademarks of IBM Corporation in the United States and other countries. Linux is a registered trademark of Linux Torvalds. dBASE is a trademark of dataBased Intelligence, Inc. Skype is a registered trademark of Skype Ltd., in the United States and other countries.

All other product names mentioned herein and throughout the entire web site are trademarks of their respective owners.

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

Permanent link:  
[http://ibexpert.com/docu/doku.php?id=01-documentation:01-12-firebird\\_interbase-character-sets:main-character-sets-in-firebird](http://ibexpert.com/docu/doku.php?id=01-documentation:01-12-firebird_interbase-character-sets:main-character-sets-in-firebird)

Last update: **2023/08/13 18:24**

