

# What is Unicode?

## An Overview

TD Core AS&DM I18N Technology (AG)

May 15, 2009

THE BEST-RUN BUSINESSES RUN SAP™



# Contents



- 1. What are Code Pages?**
- 2. What is Unicode?**
- 3. Unicode@SAP**
- 4. Thinking Unicode**

# Contents



## 1. What are Code Pages?

## 2. What is Unicode?

## 3. Unicode@SAP

## 4. Thinking Unicode

# From Characters to Code Pages



Character:	a, b, c, A, B,...
Glyph:	ä, å, å,...
Coded character:	A=65, B=66,...
Character Set:	A set of characters, to be used together (e.g. Latin alphabet)
Code page:	A set of coded characters (e.g. ISO-8859-1, Shift-JIS)
Locale:	Code page + properties and rules (e.g. isdigit, collation, ...)

# Code Page Example: ISO 8859-1



A single byte code page assigns one distinct byte to each character. The code points range from 0 to 255. Such a code pages is drawn in a 16 \* 16 matrix.

$$\begin{aligned} \text{'Ö'} &= \text{x'D6'} \\ &= 13 * 16 + 6 \end{aligned}$$

'Ö' is the character  
D6 is the code point

	!	”	#	\$	%	&	'	(	)	*	+	,	-	.	/	
0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?	
@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_	
`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	
p	q	r	s	t	u	v	w	x	y	z	{		}	~		
	ı	ϕ	£	κ	¥	ı	š	”	ø	æ	«	¬	-	ø	-	
°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿	
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï	
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß	
à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï	
ä	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ	

# Asian and Other 16-bit Code Pages



They are drawn in a 256 \* 256 matrix.

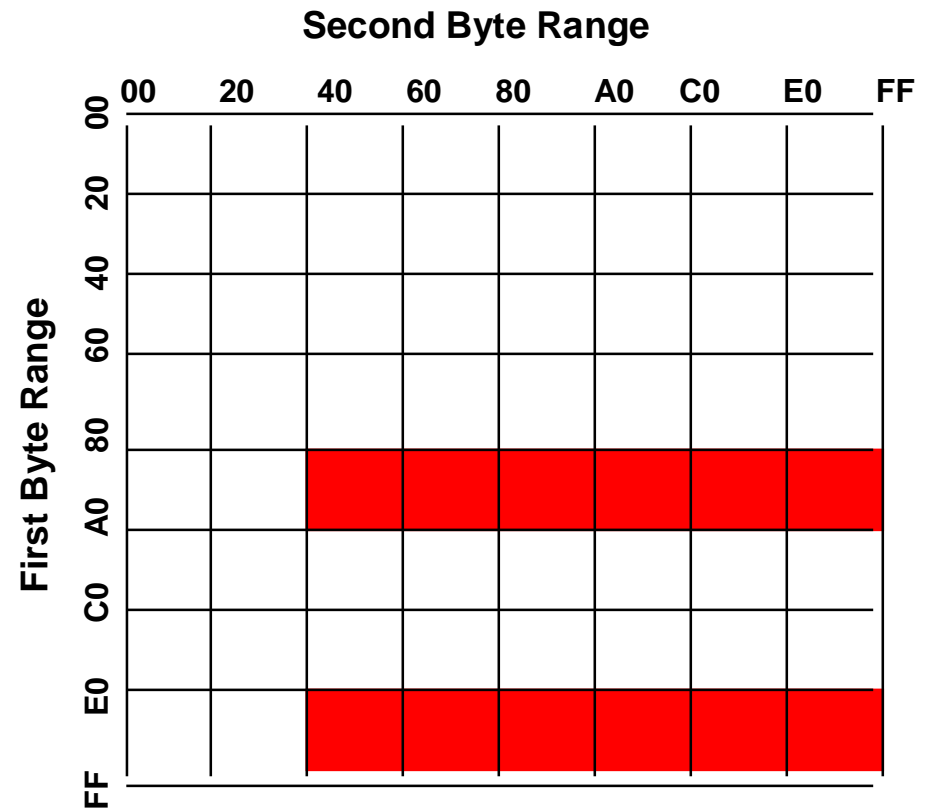
**Example:**

**The double-byte characters  
of the Shift-JIS standard**

**First byte is 80..A0 or E0..FF  
Second byte is 40..FF**

**7-Bit area 00..7F identical to ISO  
8859-x**

**Exchange of English text possible**



# Contents



1. What are Code Pages?

**2. What is Unicode?**

3. Unicode@SAP

4. Thinking Unicode

# From Code Pages to Unicode



Every standard code page supports only a certain group of languages (e.g Western European, Eastern European, Japanese)

Within one computer system only one code page can be supported in a clean way

Therefore a universal code page that supports all letters, punctuation signs, technical symbols etc. of all languages is required





Unicode is a superset of all existing character sets

Unicode encodes plain text (no rendering information)

Unicode defines characters, not glyphs (semantics, not visual representation)

Unicode unifies characters used in different scripts (CJK\* Unification)

\*Chinese, Japanese, Korean

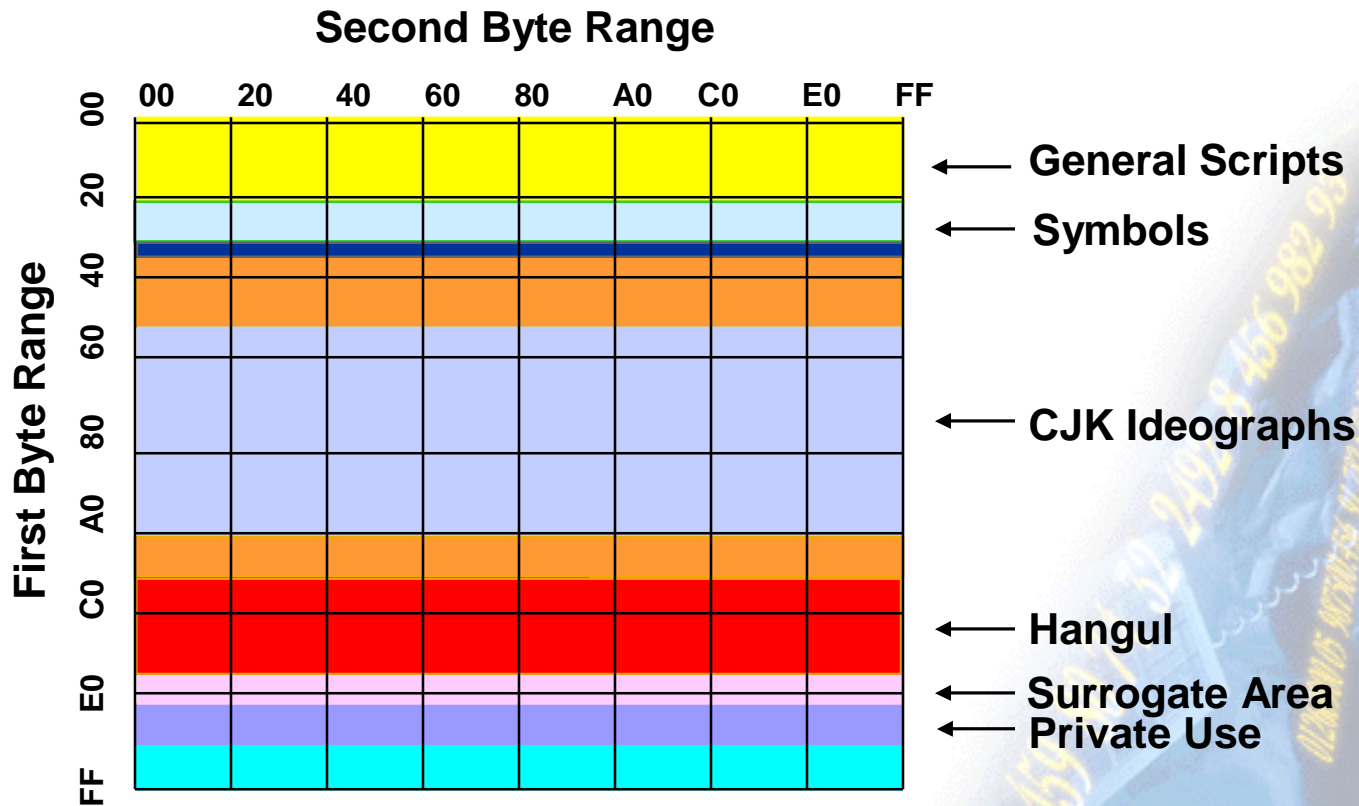
# The Unicode 'Code page'



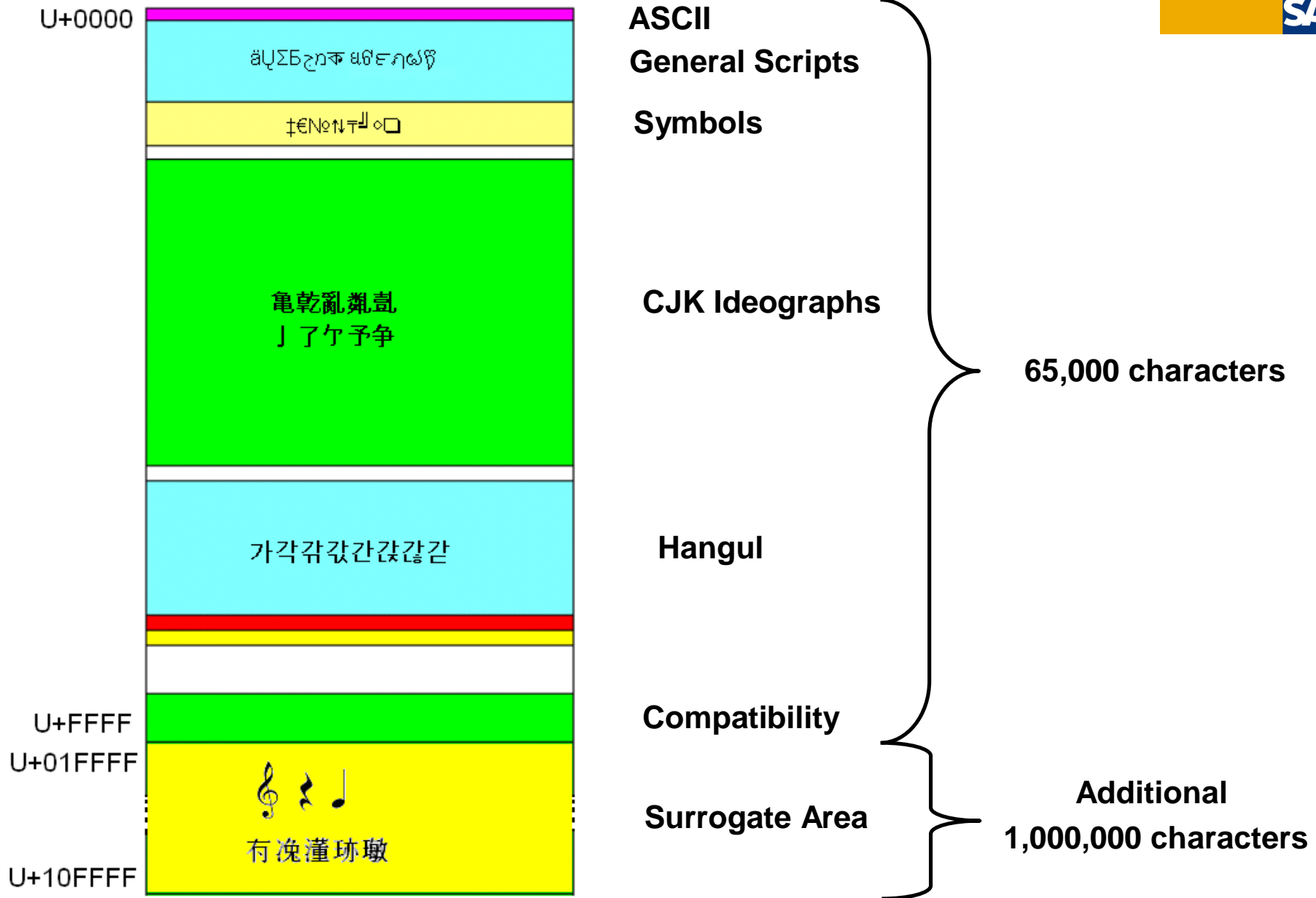
Space for 1.000.000 characters

64.000 characters coded by one 16bit code point

Further characters coded by two 16bit code points (surrogates)



# Unicode Characters



## UTF-8

- byte-based encoding scheme
- one character is coded with 1-4 bytes
- compatible with 7-bit ASCII

## UTF-16

- 16bit units
- often used characters occupy one 16bit unit
- further characters are coded with two 16bit units

## UTF-32

- 32bit units
- fixed size for all characters

All encoding forms support the same amount of characters.

# Coding example



Character	ISO-8859-1	UTF-8	UTF-16	UTF-32
A	0x41	0x41	0x0041	0x00000041
Ä	0xC4	0xC3 0x84	0x00C4	0x000000C4

# Contents



1. What are Code Pages?
2. What is Unicode?
- 3. Unicode@SAP**



## UTF-8

- for external communication (e.g file, network)
  - no endian problems
  - minimum average data size
  - limited backward compatibility to non-Unicode systems

## UTF-16

- internal (in memory)
  - best compromise between memory usage and algorithmic complexity
  - fits to Java and Microsoft environment
  - best way to migrate existing ABAP and C programs



No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.

SAP, R/3, xApps, xApp, SAP NetWeaver, Duet, SAP Business ByDesign, ByDesign, PartnerEdge and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. All other product and service names mentioned and associated logos displayed are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

The information in this document is proprietary to SAP. This document is a preliminary version and not subject to your license agreement or any other agreement with SAP. This document contains only intended strategies, developments, and functionalities of the SAP® product and is not intended to be binding upon SAP to any particular course of business, product strategy, and/or development. SAP assumes no responsibility for errors or omissions in this document. SAP does not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

SAP shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The statutory liability for personal injury and defective products is not affected. SAP has no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages

Weitergabe und Vervielfältigung dieser Publikation oder von Teilen daraus sind, zu welchem Zweck und in welcher Form auch immer, ohne die ausdrückliche schriftliche Genehmigung durch SAP AG nicht gestattet. In dieser Publikation enthaltene Informationen können ohne vorherige Ankündigung geändert werden.

Einige von der SAP AG und deren Vertriebspartnern vertriebene Softwareprodukte können Softwarekomponenten umfassen, die Eigentum anderer Softwarehersteller sind.

SAP, R/3, xApps, xApp, SAP NetWeaver, Duet, SAP Business ByDesign, ByDesign, PartnerEdge und andere in diesem Dokument erwähnte SAP-Produkte und Services sowie die dazugehörigen Logos sind Marken oder eingetragene Marken der SAP AG in Deutschland und in mehreren anderen Ländern weltweit. Alle anderen in diesem Dokument erwähnten Namen von Produkten und Services sowie die damit verbundenen Firmenlogos sind Marken der jeweiligen Unternehmen. Die Angaben im Text sind unverbindlich und dienen lediglich zu Informationszwecken. Produkte können länderspezifische Unterschiede aufweisen.

Die in diesem Dokument enthaltenen Informationen sind Eigentum von SAP. Dieses Dokument ist eine Vorabversion und unterliegt nicht Ihrer Lizenzvereinbarung oder einer anderen Vereinbarung mit SAP. Dieses Dokument enthält nur vorgesehene Strategien, Entwicklungen und Funktionen des SAP®-Produkts und ist für SAP nicht bindend, einen bestimmten Geschäftsweg, eine Produktstrategie bzw. -entwicklung einzuschlagen. SAP übernimmt keine Verantwortung für Fehler oder Auslassungen in diesen Materialien. SAP garantiert nicht die Richtigkeit oder Vollständigkeit der Informationen, Texte, Grafiken, Links oder anderer in diesen Materialien enthaltenen Elemente. Diese Publikation wird ohne jegliche Gewähr, weder ausdrücklich noch stillschweigend, bereitgestellt. Dies gilt u. a., aber nicht ausschließlich, hinsichtlich der Gewährleistung der Marktgängigkeit und der Eignung für einen bestimmten Zweck sowie für die Gewährleistung der Nichtverletzung geltenden Rechts.

SAP übernimmt keine Haftung für Schäden jeglicher Art, einschließlich und ohne Einschränkung für direkte, spezielle, indirekte oder Folgeschäden im Zusammenhang mit der Verwendung dieser Unterlagen. Diese Einschränkung gilt nicht bei Vorsatz oder grober Fahrlässigkeit.

Die gesetzliche Haftung bei Personenschäden oder die Produkthaftung bleibt unberührt. Die Informationen, auf die Sie möglicherweise über die in diesem Material enthaltenen Hotlinks zugreifen, unterliegen nicht dem Einfluss von SAP, und SAP unterstützt nicht die Nutzung von Internetseiten Dritter durch Sie und gibt keinerlei Gewährleistungen oder Zusagen über Internetseiten Dritter ab.

Alle Rechte vorbehalten.