

Interacción persona-ordenador

Internacionalización





Objetivos

- ⌘ Valorar la importancia de la internacionalización de las interfaces
- ⌘ Reconocer los problemas derivados de la traducción a otros lenguajes
- ⌘ Conocer los diferentes alfabetos y esquemas de codificación existentes
- ⌘ Saber los aspectos a tener en cuenta en la internacionalización
- ⌘ Conocer metodologías y tecnologías para internacionalizar y localizar interfaces



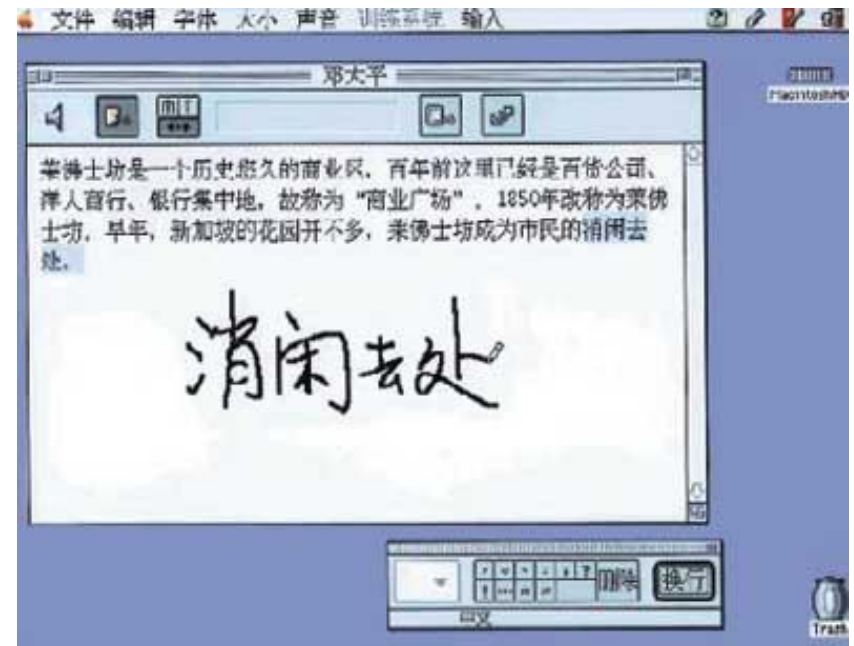
Contenidos

- ⌘ Introducción
- ⌘ Internacionalización y localización
- ⌘ Elementos de la interfaz
- ⌘ Escrituras
- ⌘ Esquemas de codificación
- ⌘ Zonas de internacionalización
- ⌘ Modelo de internacionalización y localización
- ⌘ Guía técnica



Introducción

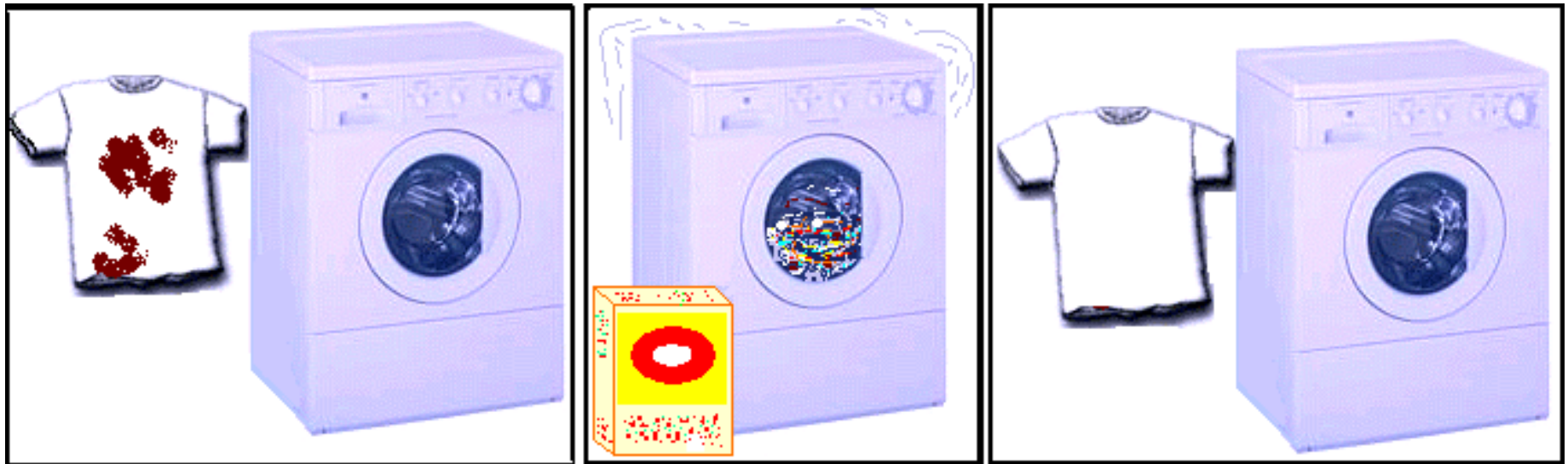
- ⌘ **Software internacionalizado**: producto que está preparado para ser utilizado fuera de la región o país donde fue creado
- ⌘ **Objetivo**: hacer llegar el producto a mercados internacionales
- ⌘ **Problema**: ajustar la interfaz a los diferentes destinos
- ⌘ **Caso extremo**: sitios web





Introducción

- ⌘ Ejemplo de problema: culturas con lenguajes que se escriben de derecha a izquierda



Copyright © 2000—2003 Tex Texin, Stéphanie Smith. All Rights Reserved



Internacionalización y localización

Internacionalización

- ⌘ Es el proceso de diseñar una aplicación de tal manera que pueda adaptarse a diferentes lenguajes y regiones sin necesidad de cambiar el código
 - ☑ Un programa internacionalizado no tiene elementos dependientes de la lengua o del contexto cultural de un país o región en el propio código
 - ☑ Los elementos textuales, como los mensajes o las etiquetas de los componentes, no están en el código, están fuera y se toman dinámicamente



Internacionalización y localización

Localización

- ⌘ Es el proceso de adaptar *software* a una región específica o lenguaje añadiéndole componentes específicos y traduciendo texto
 - ☑ Convenciones locales, cultura y lenguaje de una región particular
 - ☑ Conjunto de letras que se utilizan para escribir en una lengua (escritura)



Internacionalización y localización

Ventajas

- ⌘ El mismo ejecutable funciona en todo el mundo
- ⌘ El mercado es mayor
- ⌘ No hace falta hacer un desarrollo internacional del producto una vez acabada la primera versión
- ⌘ Se utilizan los recursos más eficientemente
- ⌘ Añadir soporte internacional a un programa acabado puede hacerlo menos estable que si se hace como una parte del proceso de desarrollo
- ⌘ El mantenimiento del código y la inclusión de nuevas localizaciones es menos costoso



Internacionalización y localización

Elementos a considerar

⌘ **Texto**

⌘ Iconos

⌘ Clips de audio

⌘ Ayuda en línea

⌘ Formato de monedas,
fechas y números

⌘ Calendarios

⌘ Medidas

⌘ Colores

⌘ Gráficos

⌘ Números de teléfono

⌘ Direcciones

⌘ Títulos honoríficos



Elementos de la interfaz

- ⌘ Cultura
- ⌘ Iconos
- ⌘ Colores
- ⌘ Calendarios, formatos y separadores de fecha y hora
- ⌘ Formatos de números y monedas
- ⌘ Ordenación
- ⌘ Unidades de medida



Elementos de la interfaz

Cultura

- ⌘ La cultura de una región o país puede hacer percibir un mismo objeto de maneras diferentes en puntos geográficos diferentes
- ⌘ Ejemplos
 - ☒ *Tan pronto como sea posible* quiere decir *inmediatamente* en Estados Unidos y *puede ser este mes* en la cultura latinoamericana
 - ☒ La mano izquierda es ofensiva en algunas culturas



Elementos de la interfaz

Cultura

- ⌘ Hay que cuidar cómo se dibujan hombres y mujeres juntos y cómo van vestidos
- ⌘ Algunos **símbolos** como la esvástica, hoz y martillo, sol naciente, cruces y estrellas representan ideas políticas y/o religiosas
- ⌘ En un país fundamentalista pueden entenderse como ofensivas cosas que en Occidente no tienen ningún significado
 - ☑ Mover la cabeza, decir adiós con la mano, mujeres en bikini, bebidas alcohólicas



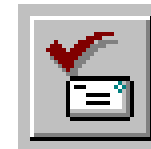
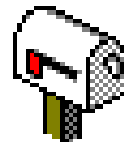
Elementos de la interfaz

Iconos

⌘ Muchos iconos presentan una dependencia más o menos importante de una determinada lengua y cultura

⌘ Ejemplos:

- ☑ Buzón de correo
- ☑ Corrector ortográfico
- ☑ Comprobación (*tick*)





Elementos de la interfaz

Colores

⌘ Tienen asociados significados a través de la tradición en la mayor parte de las culturas

⌘ Occidental

- ⊗ Negro: color asociado con el luto y los funerales
- ⊗ Blanco: asociado con el matrimonio y el nacimiento

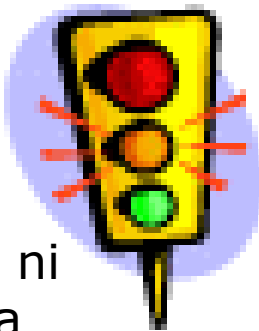
⌘ Oriental

- ⊗ Blanco: asociado con el luto
- ⊗ Rojo: asociado con el matrimonio

⌘ Cada cultura tiene sus estándares, por lo que es difícil generalizar

⌘ Ejemplo: semáforo

- ⊗ Rojo (parar), ámbar (precaución) y verde (adelante), pero ni aún estos los podemos asumir de una manera generalizada





Elementos de la interfaz

Calendarios, fecha y hora

- ⌘ Hay otros tipos de calendario aparte del Gregoriano
 - ☒ Budista, islámico, chino, hebreo, etc.
- ⌘ En el calendario gregoriano hay distintos formatos:
 - ☒ **dd/mm/aaaa** 11 de enero de 2005
 - ☒ **mm/dd/aaaa** january 11th, 2005 (países de habla inglesa)





Elementos de la interfaz

Números y monedas

⌘ Para cada país hace falta tener en cuenta el símbolo de la moneda y el formato numérico

☒ USA \$ 1,234.56

☒ Gran Bretaña £ 1,234.56

☒ Euro 1.234,56 €





Elementos de la interfaz

Ordenación

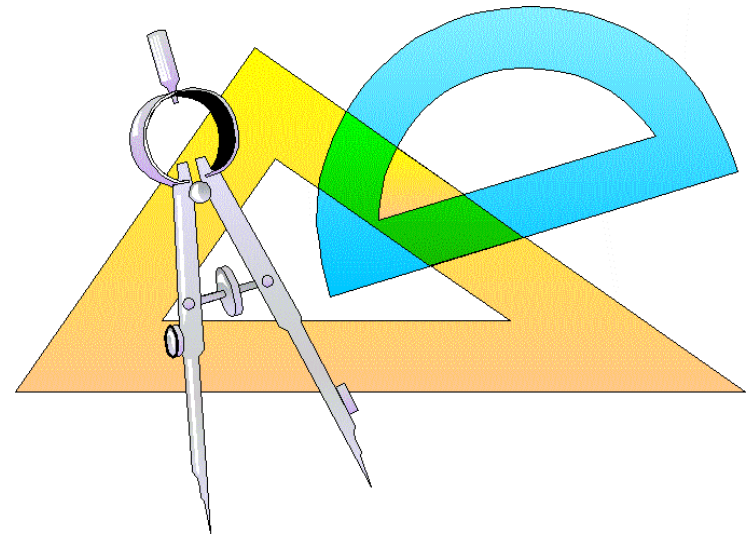
- ⌘ Las ordenaciones son importantes en el proceso de internacionalización y son un tema complejo
- ⌘ En un principio las ordenaciones de caracteres son de la A a la Z, mayúsculas primero y después las minúsculas, y los números se ordenan de 0 a 9
- ⌘ Este tipo de regla no siempre es aceptable, porque hay reglas más complicadas cuando hay que considerar un segundo carácter al mismo tiempo
 - ☒ Ejemplo de la 'ch' y la 'll' en Castellano, ya superado



Elementos de la interfaz

Unidades de medida

- ⌘ La mayor parte del mundo utiliza el sistema métrico excepto Estados Unidos
 - ☑ Utiliza la milla (1.609 m), la pulgada (2,54 cm), etc.





Escritura

- ⌘ Las escrituras se pueden clasificar en *ideogramas* y *escrituras fonéticas*
 - ☑ Un ideograma tiene un significado especial y no tiene relación con su pronunciación
 - ☑ Las letras de una escritura fonética representan determinados sonidos, como por ejemplo M
- ⌘ Los tres sistemas de escritura más importantes son:
 - ☑ Occidental
 - ☑ Oriente medio
 - ☑ Extremo Oriente





Escritura

Escrituras occidentales

- ⌘ Las escrituras occidentales son: latín, griego y cirílico
- ⌘ Características comunes a las tres escrituras:
 - ☒ Fonéticas
 - ☒ Se leen de izquierda a derecha en una línea horizontal
 - ☒ Utilizan letras mayúsculas y minúsculas diferentes
 - ☒ Utilizan numeración arábiga: 0,1,2,3,4,5,6,7,8 y9



Escritura

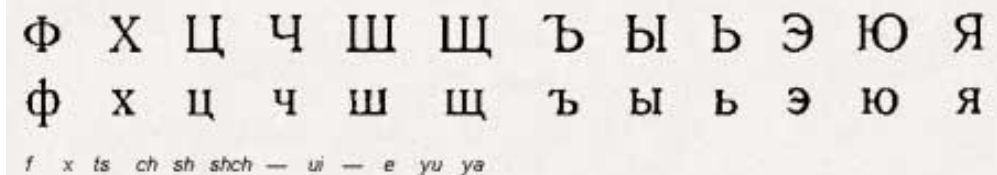
Escrituras occidentales

⌘ Escritura latina

- ☑ La más usada (casi toda Europa, parte de Asia, casi toda África, América y Oceanía)
- ☑ Proviene del latín, con letras añadidas (J, W)
- ☑ Algunas lenguas usan más letras y marcas diacríticas:
 - ☒ acento grave (` à), agudo (´ á), circunflejo (^ â), tilde (~ ã), diéresis (¨ ä), círculo (° å) y slash (/ Ø).
 - ☒ En castellano, á, é, í, ó, ú, ü, ñ, ÷, i

⌘ Escritura griega

- ☑ Usada en Grecia



⌘ Escritura cirílica

- ☑ Usada en los países eslavos (ruso, búlgaro, serbio...) y de Asia central (turcomeno, kurdo, uzbeko, kazakho, tajik...)



Escritura

Extremo Oriente

- ⌘ Están basadas en los caracteres ideográficos chinos
- ⌘ Su antigüedad se remonta a 4000 años y se utilizan de esta forma desde hace más de 2000 años
- ⌘ Actualmente se usan unos cuantos miles de caracteres

Conjunto estándar:
20.000 caracteres
Conjunto extendido:
50.000 caracteres
Mínimo para conversar:
3.000 caracteres

Carácter	Chino	Japonés	Coreano	Significado
左	tsuoh	Hidari	jua	Izquierda
右	yu	Migi	woo	Derecha
金	jin	Kin	goom	O
木	mu	moku (ki)	nahmoo	Madera
水	shui	shui (mizu)	mool	Agua
火	huo	ka (hi)	bool	Fuego
土	tu	do (tsuchi)	huk	Tierra



Esquemas de codificación

- ⌘ Cada tipo de escritura ha de ser representada en sistemas informáticos
- ⌘ Esta representación se concreta en un **esquema de codificación**
- ⌘ Estos esquemas se codifican a través de **mapas de caracteres**
 - ☑ Los mapas de caracteres permiten la representación de escrituras en el ordenador e incluyen caracteres que no están en el alfabeto como los signos de puntuación



Esquemas de codificación

Tipos de esquemas

⌘ ASCII

- ☑ Mapa de caracteres de un octeto (SBCS)
- ☑ Mapa de caracteres de doble octeto (DBCS)

⌘ Unicode



Esquemas de codificación

ASCII

- ⌘ El mapa de caracteres ASCII está basado en el alfabeto latino y consta de **128 caracteres**
 - ☒ Los caracteres en el rango de 0 a 31 y el carácter 127 son especiales, típicamente de control
- ⌘ Cada carácter es un índice en el mapa de caracteres, por ejemplo el número 65 es la letra 'A' y el 97 la 'a'
- ⌘ El código ASCII de 128 caracteres sólo se utilizaba inicialmente para el inglés
 - ☒ En principio con 128 caracteres había bastante
 - ☒ Al ampliarlo a nuevos idiomas se empezaron a utilizar los restantes 128 caracteres



ASCII

Mapa de un sólo octeto (SBCS)

- ⌘ Son mapas de caracteres de hasta 256 caracteres ($256=2^8$)
- ⌘ Son una extensión del código ASCII, al que se han añadido 128 caracteres, del 128 al 255 (mapa de caracteres extendido)
- ⌘ ISO
 - ☒ 8859-2 Latín europa del este
 - ☒ 8859-5 Cirílico Europa del este
 - ☒ 8859-1 Latín Europa del Oeste
 - ☒ 8859-7 Griego
 - ☒ 8859-3 Turco
 - ☒ 8859-8 Hebreo
 - ☒ 8859-6 Árabe



Mapa de un sólo octeto

ASCII ISO 8859-1 (Latin-1)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
20		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
80			,	f	„	...	†	‡	^	%	Š	<	Œ			
90		‘	’	“	”	•	—	—	~	™	š	>	œ			ÿ
A0		ı	€	£	¤	¥		§	"	©	ª	«	¬	-	®	¯
B0	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï



Mapa de un sólo octeto

ASCII ISO 8879

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
20		!	∇	#	∃	%	&	⋈	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	≅	A	B	X	Δ	E	Φ	Γ	H	I	∅	K	Λ	M	N	O
50	Π	Θ	P	Σ	T	Υ	ς	Ω	Ξ	Ψ	Z	[∴]	⊥	_
60	—	α	β	χ	δ	ε	φ	γ	η	ι	φ	κ	λ	μ	ν	ο
70	π	θ	ρ	σ	τ	υ	ω	ω	ξ	ψ	ζ	{		}	~	
80																
90																
A0		T	'	≤	/	∞	f	♣	♦	♥	♠	↔	←	↑	→	↓
B0	°	±	"	≥		∞	∂	•		≠	≡	≈	...		—	↵
C0	ℵ	℥	℔	℘	⊗	⊕	⊙	∩	∪	⊃	⊇	⊄	⊂	⊆	∈	∉
D0	∠	∇			™	Π	√	·		^	∨	↔	←	↑	⇒	↓
E0	◊	<				Σ	∫		∫	∫		∫	∫	{	∫	
F0		>	∫	∫		∫	∫		∫	∫		∫	∫	}	∫	



ASCII

Mapa de doble octeto (DBCS)

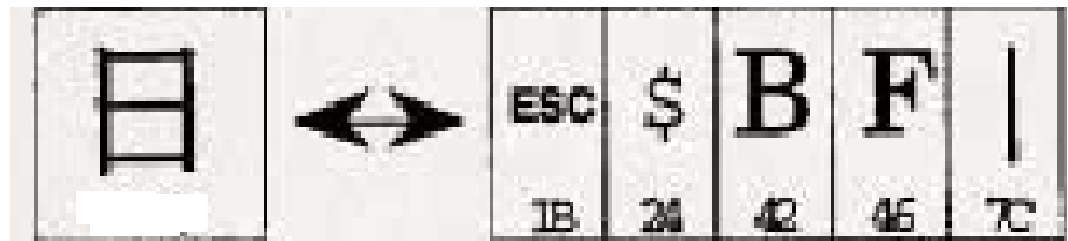
- ⌘ Las escrituras del extremo oriente utilizan caracteres ideográficos
- ⌘ Necesitan un mapa de caracteres mayor de 256 caracteres
- ⌘ Para resolverlo se creó el mapa de caracteres de doble octeto
- ⌘ Notación de 16 bits: 65.536 (2^{16}) caracteres
- ⌘ El código DBCS contiene **caracteres de 1 y 2 octetos**
 - ☒ Los de un sólo octeto se corresponden con el código ASCII
 - ☒ En el doble octeto algunos caracteres se definen como octetos de comienzo y el segundo como octeto de código



Mapa de doble octeto

Código KANJI

- ⌘ Definido por el JIS (*Japan Industry Standard*)
 - ☑ El JIS nivel 1 contiene unos 3.000 caracteres corrientes
 - ☑ El nivel 2 añade 3.500 caracteres más
 - ☑ El tercer nivel se ha definido recientemente e incrementa el mapa de caracteres en más de 6.000 caracteres
- ⌘ JIS se basa en un formato de 7 bits que requiere secuencias de ESCAPE especiales para entrar o salir del modo DBCS





Esquemas de codificación

Problemas

- ⌘ **Duplicidad** en la codificación de caracteres
- ⌘ Múltiples códigos de caracteres **inconsistentes** debido al conflicto entre estándares nacionales y de la industria
- ⌘ El código ASCII de 7 bits o el de 8 bits están limitados a 128 y 256 posiciones de código y son **inadecuados** en un entorno global
- ⌘ **Internet** ha añadido un punto más a la demanda de un conjunto de caracteres único en el mundo



Esquemas de codificación

Unicode

⌘ **Consensuado**, participan

☑ Apple, Borland, Digital, Hewlett-Packard, IBM, Lotus, Metaphor, Microsoft, Next, Novell, Research Libraries Group, Sun, WordPerfect, Xerox

- ⌘ Es un esquema de codificación uniforme y de tamaño fijo para todo tipo de caracteres de las escrituras más importantes del mundo, así como símbolos técnicos
- ⌘ Trata los caracteres alfabéticos, ideográficos y los símbolos de la misma manera, de forma que se pueden mezclar



Esquemas de codificación

Unicode

- ⌘ Es compatible con el código ASCII
- ⌘ Utiliza una codificación de 16 bits
- ⌘ No necesita secuencias de ESCAPE

ASCII/8859-1 Text

A	0100 0001
S	0101 0011
C	0100 0011
I	0100 1001
I	0100 1001
/	0010 1111
8	0011 1000
8	0011 1000
5	0011 0101
9	0011 1001
-	0010 1101
l	0011 0001
	0010 0000
t	0111 0100
e	0110 0101
x	0111 1000
t	0111 0100

Unicode Text

A	0000 0000 0100 0001
S	0000 0000 0101 0011
C	0000 0000 0100 0011
I	0000 0000 0100 1001
I	0000 0000 0100 1001
	0000 0000 0010 0000
天	0101 1001 0010 1001
地	0101 0111 0011 0000
	0000 0000 0010 0000
س	0000 0110 0011 0011
ل	0000 0110 0100 0100
ا	0000 0110 0011 0111
م	0000 0110 0100 0101
	0000 0000 0010 0000
α	0000 0011 1011 0001
κ	0010 0010 0111 0000
γ	0000 0011 1011 0011



Esquemas de codificación Unicode

- ⌘ En resumen, Unicode es:
 - ☑ Universal (capacidad amplia)
 - ☑ Uniforme (anchura fija)
 - ☑ No ambiguo
- ⌘ Unicode proporciona códigos para más de 39.000 caracteres de las lenguas más importantes (quedan cerca de 18.000 sin usar y 6.400 reservados para uso privado de las aplicaciones)

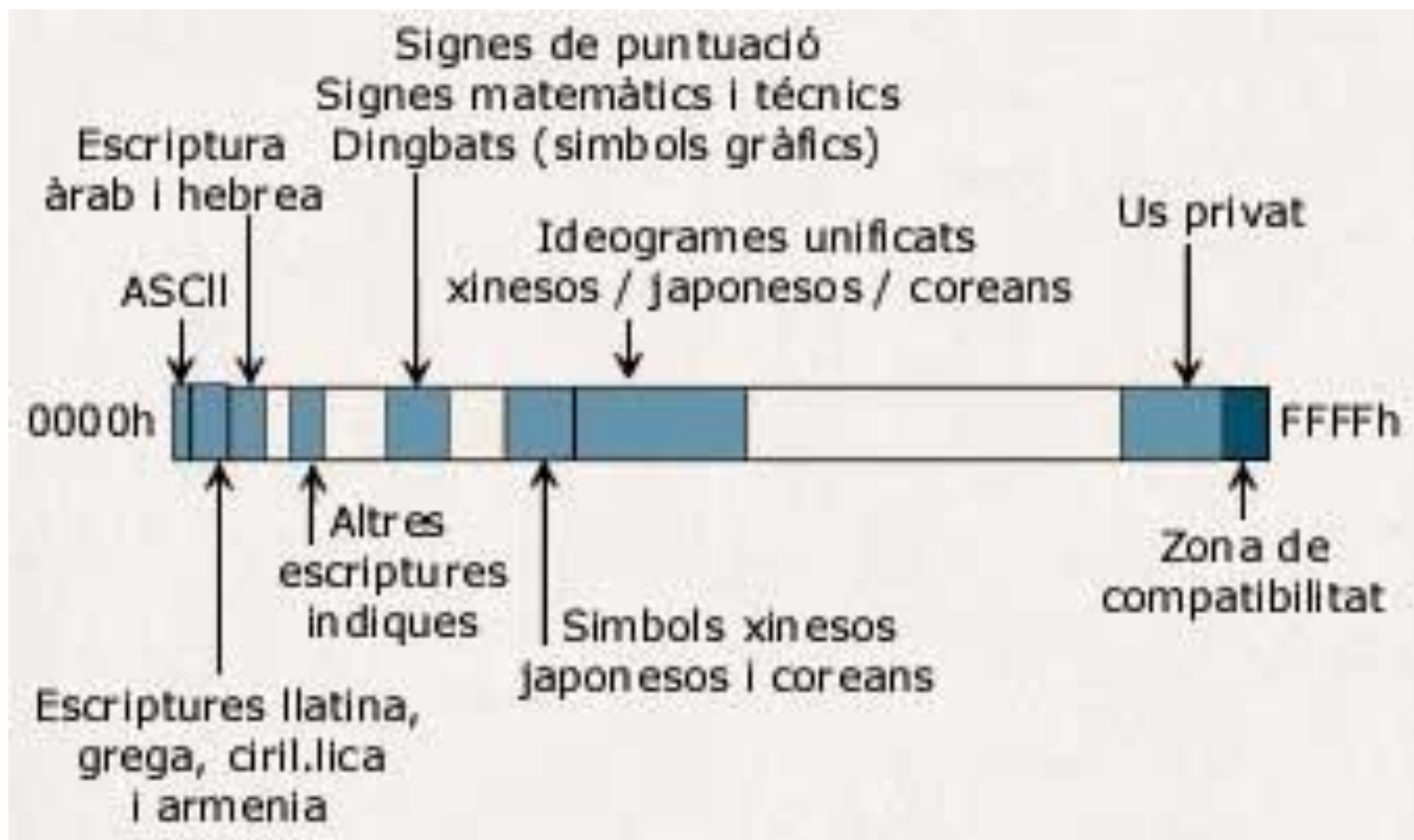
Unicode		2022 + 8859 + JIS
A 0041	↔	A 41
å 00E5	↔	å E5
ı 0645	↔	ESC - G å 1B 2D 47 E5
£ 03B5	↔	ESC - F å 1B 2D 46 E5
ı 0131	↔ or	ESC - C İ 1B 2D 43 B9 ESC - I ý 1B 2D 49 FD
☐ 65E5	↔	ESC \$ B F 1B 24 42 46 7C



Esquemas de codificación

Unicode

⌘ Disposición de símbolos y letras





Zonas de internacionalización

Zona	Idiomas	Conjunto de caracteres	Escritura	Direccionalidad del texto
Europea	Europa, oeste, central y este, griego, ruso, turco, indonesio	Un byte	Latín, griego, Cirílico	Izquierda a derecha
Oriente medio	Arabe, Hebreo	Un byte	Árabe, Hebreo, Latín	Bidireccional
Extremo oriente	Chino tradicional, Chino simplificado, Japonés, Coreano	Multibyte	Kana, hangul, caracteres ideográficos	Horizontal y vertical
Thai	Thai	Un byte	Thai	Izquierda a derecha



Modelo de localización

⌘ Esquema de un producto localizado:

Recursos
localizados

+

Código de la
aplicación

=

Producto
localizado



Modelo de localización

⌘ Esquema de un producto localizado:



- ⌘ Contiene datos pero no código
- ⌘ Contiene cadenas de caracteres y gráficos utilizados en la interfaz
- ⌘ Contiene elementos específicos de la localización realizada



Modelo de localización

⌘ Esquema de un producto localizado:



⌘ Contiene el código de la aplicación que se ejecuta en todos los países e idiomas



Modelo de localización

⌘ Claves para el diseño de un producto localizado:

- ☑ Extraer todos los datos, textos y elementos gráficos y colocarlos en un fichero de datos separado
- ☑ Organizar y documentar los ficheros
- ☑ Crear un proceso para construir el producto localizado



Guía técnica

- ⌘ Veamos algunos problemas técnicos generales en la planificación de la localización



Guía técnica

Puntos a revisar

- ⌘ Elementos específicos de la localización
- ⌘ Convenciones en el procesamiento de texto:
 - ☑ ordenación, clasificación de caracteres, pronunciación, guionado, funciones de gestión de cadenas e intercambio de ficheros (importación y exportación de ficheros)
- ⌘ Gestión de las entradas
 - ☑ soporte para *drivers* de teclado y otros métodos de entrada
- ⌘ Gestión de las salidas



Guía técnica

Traducción del texto

- ⌘ Todas las lenguas tienen gramática y sintaxis diferentes
- ⌘ El texto traducido cambia de tamaño, hay que prever que pueda crecer
- ⌘ *Ejemplos:*
 - ☒ *Mover* se traduce como *verschieben* en alemán (5 contra 11)



Guía técnica

Aspectos a tener en cuenta

- ⌘ Mensajes
- ⌘ Gramática y sintaxis
- ⌘ Diseño de menús
- ⌘ *Prompts*
- ⌘ Barras de estado
- ⌘ Cajas de diálogo
- ⌘ Iconos



Guía técnica

Organización de diálogos

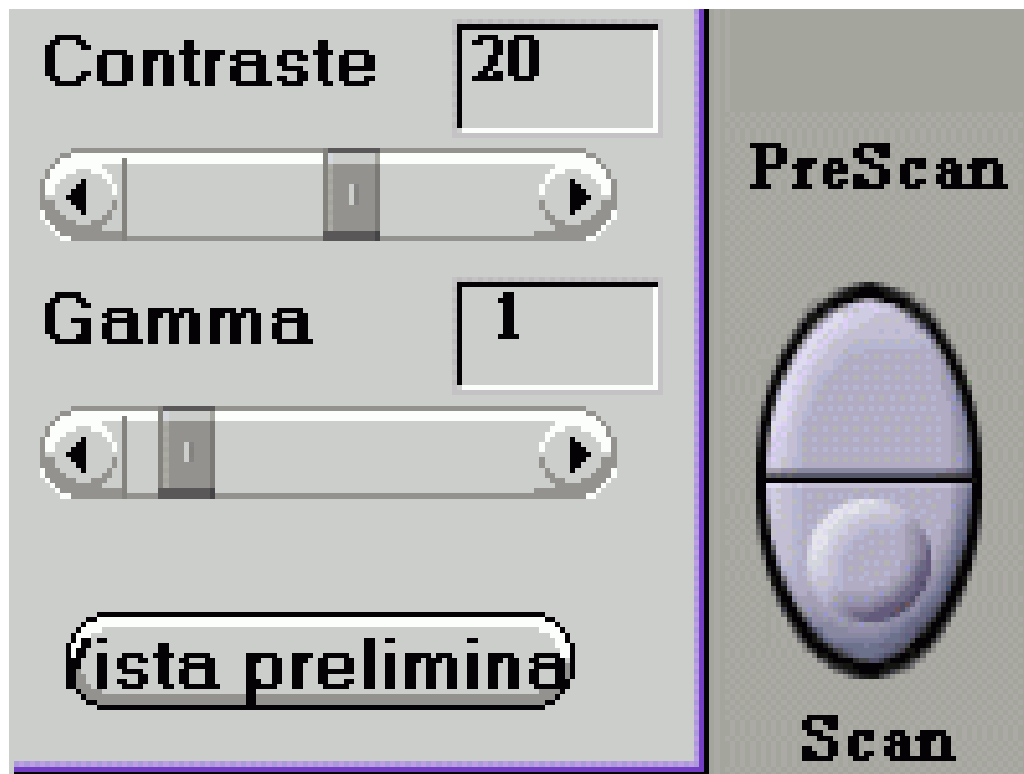
- ⌘ La organización de controles y elementos en una ventana ha de considerar la expansión de las etiquetas de texto
- ⌘ En hebreo y árabe la información se escribe de derecha a izquierda (hace falta invertir la presentación)
- ⌘ Algunos idiomas incluyen marcas diacríticas que requieren espacio adicional
- ⌘ Regla: permitir un 30% de expansión



Guía técnica

Organización de diálogos

⌘ Ejemplo:



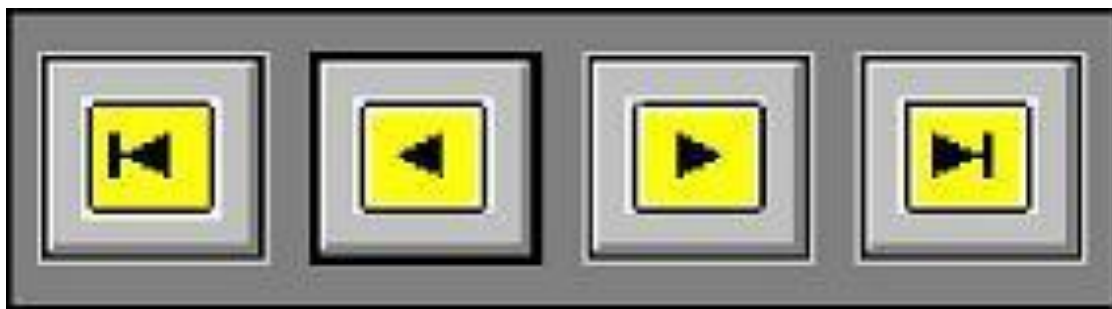
Escáner Mustek



Guía técnica

Organización de diálogos

⌘ Ejemplo:



*Controles de
reproducción*

primero

anterior

siguiente

último

Izquierda a derecha

último

siguiente

anterior

primero

Derecha a izquierda



Guía técnica

Barra de estado

- ⌘ No se ha de llenar la barra de estado con texto
- ⌘ La versión inglesa sólo ha de ocupar el 50% de la barra de estado



Guía técnica

Menús

⌘ Evitar menús muy densos

- ☑ Límite 80 caracteres

⌘ Las teclas aceleradoras coinciden con el primer carácter de una palabra. Esto puede dar problemas al traducir. Habrá que permitir que la tecla aceleradora sea otra distinta de la inicial



Guía técnica

Elementos culturales

⌘ Dirección postal

Name:

Address:

City: State: Zip:

Dirección típica en EEUU



Guía técnica

Elementos culturales

⌘ Dirección postal - solución

Name:	<input type="text"/>
Email:	<input type="text"/>
Address1:	<input type="text"/>
Address2:	<input type="text"/>
Address3:	<input type="text"/>

Dirección flexible



Guía técnica

Elementos culturales

⌘ Objetos cotidianos



*¿Qué objeto
elegiremos si el
perro nos pide
una pelota?*



Guía técnica

Desarrollo

- ⌘ Documentar todos los recursos que tienen que ser localizados (y los que no)
- ⌘ Es útil hacer dos localizaciones en paralelo para poder evaluar la internacionalización
- ⌘ Es importante disponer de herramientas y utilidades para hacer la localización



Guía técnica

Ejemplo: lenguaje Java

⌘ Soporta Unicode:

- ☑ Tiene un tipo *char* de 16 bits (en C++ depende de la plataforma)
- ☑ Posee una notación para definir un carácter Unicode constante: `'\uxxxx'`

⌘ Posee objetos predefinidos

- ☑ *Class java.util.Locale extends Object*. Un objeto *Locale* representa un identificador de una región



Conclusiones

- ⌘ El mercado informático es global y para competir es preciso hacer programas para clientes de todo el mundo
- ⌘ La internacionalización es una **necesidad**, no una opción
- ⌘ Es preciso tenerla en cuenta **desde el principio** del diseño siguiendo una metodología adecuada
 - ☐ Ejemplo de Microsoft