CTT534 – Thiết Kế Giao Diện
HK II 2013 – 2014

Internationalization (i18n)

26/5/2014

Slides adapted from MIT CS Course
6.813/6.831

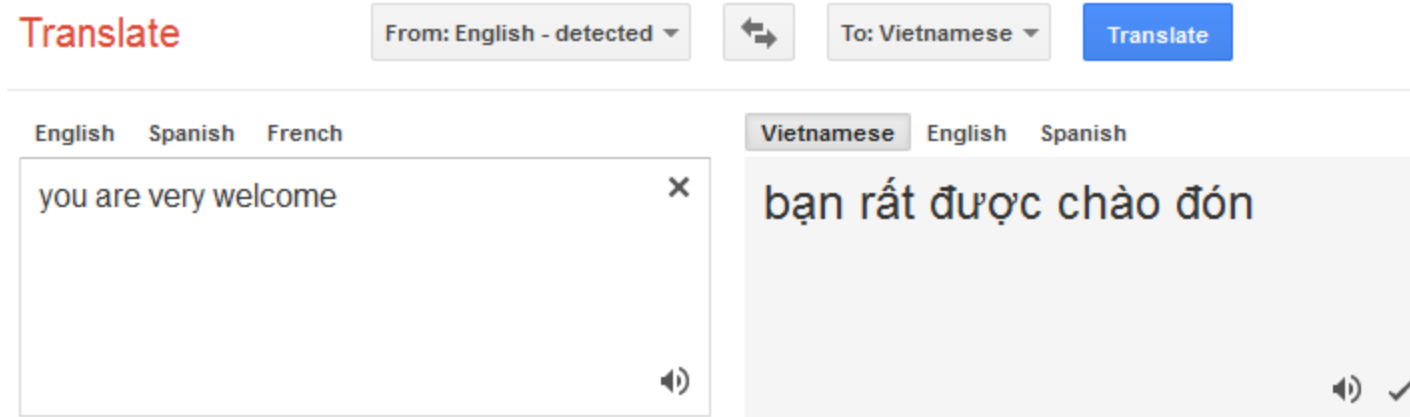# Internationalization and localization

- **Internationalization**
  - The process of making software ready for translation or adaptation into different languages and cultures
  - Often called as i18n – 18 characters between 'i' and 'n' in the word "internationalization'

- **Localization**
  - The process making software to a specific language and culture
  - More than just translate, find, and replace words

Text translation

- **All user-visible text has to be translated, including error messages**
  - Text in code
  - Text in images
  - Text in data files
- **Translation results in the change of text length**

Translate From: English - detected ▼ ⇄ To: Vietnamese ▼ **Translate**

English Spanish French

you are very welcome ✕

🔊

Vietnamese English Spanish

bạn rất được chào đón

🔊 ✓

Problem of translation

Different scripts

- English
    - This is the class on user interface design
- Vietnamese
    - Đây là lớp học về thiết kế giao diện người dùng
- Russian
    - это класс дизайна пользовательского интерфейса
- Chinese (traditional)
    - 這是用戶界面設計類
- Arabic
    - هذه هي الطبقة في تصميم واجهة المستخدم

Text direction

- Left-to-right
  - English, Vietnamese, French, etc.
- Right-to-left
  - Arabic, Hebrew
- To-down, left-to-right, right-to-left
  - Chinese, Korean, Japanese

- Text direction affects drawing, screen layout, editing, etc.

حكم قضائي يبعد أحمد شفيق مجدّداً عن سباق الرئاسة في مصر
البيان – قبل 1 ساعة

زوجة بيل غيتس تربي أولادها على التقشف
NEW الديا8... – قبل 4 ساعات

أزمة جديدة بين القاهرة والرياض.. ملعبها مدينة جدة السعودية
NEW الديا8... – 2012/05/08

تركيا ترفض تسليم طارق الهاشمي
روسيا اليوم – قبل 1 ساعة

ديكا: أوباما استغل عميلا سعوديًا في دعايته الانتخابية
الوفد – قبل 2 ساعات

منى أمرشا: شذى حسون قلبها "أسود"
الوطن – قبل 5 ساعة

هكذا ستطلّ هيفا وهبي في مسلسل "مولد وصاحبه غايب".. بالصور
NEW الديا8... – قبل 5 ساعة

تأجيل محاكمة عز في قضية الدخيلة إلى الغد
الوفد – قبل 4 ساعات

مصدر مسئول: تشكيل فرق بحث لتحديد أماكن المتهمين في أحداث العباسية
الشروق – قبل 3 ساعات

غادة عبد الرازق تجسد شخصية "الشيخة موزة"
العرب اونلاين – قبل 4 ساعات

Sort order

- Unicode is often used
- It is not always ordered according to a language
- Upper/lowercase affects order
- Java's String.compareTo() does not work as expected all the times

U+	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
0010	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
0020	SP	!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
0030	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0040	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0050	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
0060	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0070	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL
0080	PAD	HOP	BPH	NBH	IND	NEL	SSA	ESA	HTS	HTJ	VTS	PLD	PLU	RI	SS2	SS3
0090	DCS	PU1	PU2	STS	CCH	MW	SPA	EPA	SOS	SGCI	SCI	CSI	ST	OSC	PM	APC
00A0	NBSP	¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	SHY	®	¯
00B0	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
00C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É		Ë	Ì	Í	Î	Ï
00D0	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
00E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
00F0	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Formatting

- Numbers
  - US/UK: 1,000.5 (one thousand point five)
  - Vietnam, Germany: 1.000,5
  - France: 1 000,5
- Date
  - US: 5/10/2012   (M/D/Y)
  - VN and others: 10/5/2012   (D/M/Y)
- Time
  - US: 12:30 PM (H:M)
  - VN: 12.30 chiều

Color conventions

- White
  - ❑ East Asia (China, Vietnam, etc.): death, used in funeral

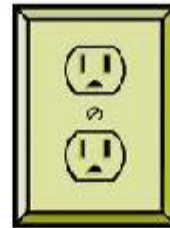 - ❑ US: purity, used in wedding

- Red
  - ❑ East Asia (China, Vietnam, etc.): luck, on new year celebrations

  - ❑ US: sometimes red refers to danger, e.g. stop sign

Icons

- Common icons in one culture may not be common in others

Guidelines on supporting I18N

- Message and resource files
- Unicode
- Bi-directionality
- Formatting libraries
- Separating processing and structure from presentation

Message and resource files

- Message/resource files are separated from source code
  - Storing messages, images, icons
  - Can be translated and created without changing code
- Pay attention to messages with dynamic parts
  - E.g., "*<N>* users have visited since *<date>*"
  - Above is the structure in English, but in Vietnamese, it's different

    "Số người đến thăm từ ngày **<date>** là **<N>**"

Unicode

- Many character sets have been used
  - ASCII
  - ASCII extension
  - Unicode
- Use Unicode for better i18n support

Bi-directionality

- Bi-directional text display and editing
  - Storing in memory may be different from displaying
    - String in memory: "This is **arabic text**"
    - Shown on screen:
      - English: "This is **txet icbara**"
      - Arabic: "**txet icbara** This is"
- Use UI toolkits provided in the language

Formatting libraries

- Use libraries provided in the language/framework
  - Java: http://java.sun.com/javase/technologies/core/basic/intl/
  - MS .NET: http://msdn.microsoft.com/en-us/goglobal/bb688096

Separating processing from presentation

- Use and store replaceable images and icons
    - Separate them from code
- Separate and allow configuring fonts and colors without changing code

Summary

- **Benefits of supporting internationalization**
    - Writing once, deploying in many languages and cultures
    - Avoid changes in code → reduce risks and cost
- **Internationalization makes a user interface ready for localizing into different locales**
- **We should consider differences in**
    - Languages
    - Scripts
    - Cultures
    - Formatting conventions