Software Usability

Course notes for CSI 5122 - University of Ottawa

2021 Deck G:

Internationalization

Timothy C. Lethbridge

< Timothy.Lethbridge@uottawa.ca >

http://www.eecs.uottawa.ca/~tcl/csi5122

Basic terminology

Locale

- Set of features that can be varied depending on the language and culture of the user or the data

Internationalization (I18N)

- The process of designing software so that it can be easily adapted to different locales

Localization (L10N)

- The process of adapting software to a locale

Different aspects of locale

The following can be treated somewhat separately

- The user's preferred locale
  - —E.g. formats for dates, times etc.
- The language of the UI
  - —The system might not have a language corresponding to the user's preferred locale
- The locale of the data
  - —e.g. currencies, formats embedded in it

Names and Titles

Some countries require you to specify Mr, Dr, Eng. Etc.)

- These titles do not necessarily translate

The family name is not always last

You do not always sort based on the family name

- In Iceland you sort based on ‘first’ name

Salutations in letters (e.g. Dear) are different in different locales

Calendars

The Gregorian calendar should not always be assumed

- Proper localization of some software requires the use (at least as an option) of calendars distinct to a culture
  - E.g. emperor-era calendar in Japan
  - Calendars of various religions where year 0 was not 2021 years ago

- Fiscal-year based calendars vary widely
  - Some have 13 months (364/28) or 53 weeks

Humour

Generally does <u>not</u> translate

- Puns are language-specific
- People are sensitive to different things in different cultures
  - —Jokes/cartoons can be offensive

Icons

Icons that are a play on words do not translate

- E.g.
  - —A tray for a server application
  - —A rocket for launching an application
  - —A running person for running an application
  - **—“B”**, “*I*”, “<u>U</u>”

Icons … continued

‘$’ does not mean ‘money’, but means ‘dollar’

- In many contexts it implicitly means ‘American dollar’

Some concepts have been found extremely hard to represent as an icon

- E.g. Sorting (‘A->Z’ is not universal)

Images of people or body parts such as hands

- Considered inappropriate in some cultures
- What skin colour do you use?

Language selection

Avoid using national flags from which people pick their preferred language

- Multiple countries use the same language

What order do you display languages?

What language do you display languages

- In the language itself
- With a translation in the language of the operating system

Oral pronunciation

Important for voice I/O systems

- Don’t forget to take pronunciation/accent into account
- Higher recognition accuracy can be obtained by tailoring voice input to regional dialects
- Voice output in the wrong dialect can make an application sound ‘foreign’

Capitalization

Some lowercase characters have different uppercase equivalents in different locales

- E.g.’ ı’ becomes ‘I’ in Turkish, whereas ‘i’ is capitalized with a dot on top.

There is no such thing as UPPERCASE for many languages

Punctuation

'!', '?' and '#' are not consistently used among languages

—In Spanish: ¿ … ?

— '#' does not mean 'number'

—In French, a space precedes a ?

Use of '/' can be confusing

—Swap rows/columns/filters

—Show/hide display cues

—Page 1/2 vs. 1/2 page

Cultural references

Common problems:

- Normal business hours / business days
- Ways payments are made
  - —Some countries still require/allow use of a PIN on a credit card
  - —Payment methods vary considerably (WeChat / Apple / Samsung / Google, etc.)
- Different styles of addresses

Language ≠ Culture

English products are sold in more countries than translated products

- Many countries (e.g. in Africa, India) have too many different languages and accept English software

Language ≠ Culture (continued)

A Norwegian user:

- May not find a product with a UI in his/her language, so will accept an English or Swedish one

- But will want the software to work with Norwegian *data*:
  - —Currency
  - —Language

Date formats

Date separators depend on locale

- ‘/’ , ‘-’ , ‘.’

Variables in document templates:

- <date> <time> <filename>

‘am’ and ‘pm’

- Not universally used (many cultures use 24 hour clock)

Date formats continued

ISO standard dates are unambiguous

- yyyy-mm-dd hh:mm:ss

Non ISO date 01-03-02 means different things in different locales.

- If not using ISO, then display dates in the locale of the user
- Preferably use a ‘long’ form with the month spelled out (in the correct language)
- Spell out day of week (‘Mon’) to help prevent errors
- However, the UI might not have been translated into the local language
  - Use the spelled-out date in the local language anyway

Numeric formats

Depends on locale, not language of application

Group separator

- Number of digits in a group
  - —In English and ISO it is 3

- Group separator
  - —In English ‘,’ , but ISO uses space, and some locales use ‘.’ or none

- Do you use the group separator for 1000?

Numeric formats (continued)

Decimal separator

- ‘.’, ‘·’, ‘,’

Negative symbol

- ‘-’, ‘~’, ‘(…)’
- Can be positioned before or after the point
- May require a space between the symbol and the number

Currency

Use the currency symbol of the data!

- I.e. $ doesn’t automatically translate to £ or € when the locale changes

Format depends on the user’ s locale, not the currency

- Differences in formats:
  - —Symbol
  - —Position (before or after the currency)
  - —Blanks separating the symbol from the data

Currency, continued

Different ways of expressing US$1000

- —$1000 (In the US, or in Canada and the UK if the application doesn't mix currencies)
- —US$1000 (In English Canada, if the application mixes currencies)
- —1000 $ (In most French locales)
- —1000 USD when mixing large numbers of currencies

Strong currencies need decimal precision (e.g. 2 digits after the decimal point for cents)

Currency, continued

You may have to display all data in two currencies in some locales

Summing payments made over a period of time

- Beware that different exchange rates will have been in effect
- Many complex rules to do this that are highly variable

Paper size

‘Letter’ in most of the Americas; ‘A4’ everywhere else

- Does not depend on language

Poses distinct problems for generating printouts and pdf files

- Make sure your output can fit on both paper sizes

Measurements

Be aware of the need to use imperial or metric units

- —Consider user preferences
- —But also understand industrial norms
  - Even in the US, many industries are metric

Beware of odd measurements in data

- —You may not want people working with multiples of 2.54cm or 0.3937 inch

Watch out for precision loss due to repeatedly converting

Addresses

Don' t rely on a fixed number of lines

Don' t rely on a particular order of address elements

- E.g. Street, City, Province, postal code is *not* universal
- E.g. Postal code in Canada comes after the province, but in many European countries it comes before the city

Addresses, continued

What language should an address be written when sending mail?

- The language of the *destination*
- Except that the <u>country</u> should be written in the language of *origin*

Phone numbers

Dependent on the region of the number, not on the user' s locale

- Except for the need to add an international dialing code

Numbers and number formats change over time

Phone numbers, continued

Allow for free-format numbers

- Keep them in the way the user entered them
- Allow the user to enter them free-form, including characters such as
- Allow for extensions in numbers
- Edit numbers automatically to meet needed local format

Free (1-800) numbers are not international

- Although there are also some new international free numbers
  - —From Canada dial 011 800

Sorting

English	German	Swedish
aA	aAäÄ	aA
bB	bB	bB
cC	cC	cC
dD	dD	dD
eE	eE	eE
fF	fF	fF
gG	gG	gG
hH	hH	hH
II	II	II
jJ	jJ	jJ
kK	kK	kK
lL	lL	lL
mM	mM	mM
nN	nN	nN
oO	oOöÖ	oO
pP	pP	pP
qQ	qQ	qQ
rR	rR	rR
sS	sS	sS
tT	ß	tT
uU	tT	uU
vV	uUüÜ	vV
wW	vV	wW
xX	wW	xX
yY	xX	yYüÜ
zZ	yY	zZ
	zZ	åÅ
		äÄ
		öÖ

Translatability

If a string can be viewed by a user, it must be translatable!

Concatenations

- Due to gender and number agreement, as well as the standard of order in a sentence
- E.g. Page number -> Numéro de page
- E.g. Number of pages -> Nombre de pages

Translatability ...

Expansion of text

- Many other languages can take at least 30% more space
  - —Allow for this, or else the UI may have to be redesigned

- Narrow columns often cannot accommodate long German words

Translatability ...

- The more compact the English writing, the longer the translation
  - 'Telegraphic' style does not translate well

- Abbreviations may have to be expanded when translated
  - E.g. 'QTD' is common in financial applications (Quarter to date)
  - (Trimestre corrent fino ad oggi) (Italian)

Translatability ….

Ambiguous phrases

- How would a translator translate the following menu items?
  - ‘Display options’
    - Options of the display
    - Show the options (all of them)
  - ‘Update version’
    - Change to the new version
    - Show the current version

Expert English users will often understand these in context

Translatability …

When you give text to translators, make sure they know for each piece of text

- E.g.. a menu label, menu item, group box etc

… the purpose

… the part of speech

- Noun, verb etc.
  - All items in a menu or set of check boxes should have the same grammatical structure

Design of internationalized software

Create a *resource file* for each locale and language

- All strings to be displayed (except data) are taken from this file
- English is just one language

Decisions about languages and character sets need to be made early in design

Design of internationalized software

Special care must be taken when integrating 3rd party software

- May not follow the same internationalization standards as you want

Memory required by the application may vary according to the language used

Scripts, fonts and character sets

Definitions:

- *Code point*: Number representing a character
- *Glyph*: Visual appearance of a character
- *Extended character*: Anything with a code point > 128

Definitions, continued

- *Special character*: A term considered a bit derogatory
- *Accented character*: Character whose glyph incorporates an accent
  - —as opposed to having an accent added when displayed
- *Diacritic*: A symbol used to modify the appearance of characters
  - —E.g. the cedilla (ç) is a diacritic, not an accent

Complex scripts

Scripts with many diacritics and character shapes

- E.g. In Arabic, characters look different depending on their position relative to others
- E.g. in Thai, diacritics can be stacked on top of each other several levels
- Also in Thai, spaces separate syllables, not words
  - ‘ABCD’ ‘AB CD’ ‘A BCD’ mean different things, causing problems at line breaks

Scripts that do not run left-right

E.g. Arabic

Mirror the UI. Everything on left moves to the right etc.

- But watch out for images etc.
- Problem if the text says, the diagram in the top-right corner

Text entered right-left

- But numbers may still be entered left-right

Some languages run top-bottom

Large ideographic scripts

E.g. Japanese, Chinese

Many standards and vendor-specific implementations

Use multiple bytes for each character

- Standard C functions 9e.g. strncpy) do not work properly and can chop off parts of characters

Inter-line spacing must be larger than Latin fonts since the characters are ‘taller’

Miscellaneous problems with multilingual software

Inability to enter needed text at a keyboard!

Upper-casing is absent or different in different languages

—Some uppercasing algorithms will translate text into garbage

Open French text on a Chinese operating system:

—Extended characters are displayed as Chinese characters and subsequent characters disappear!

Unicode

Intended to display all characters in all languages

- Including technical symbols

Allows exchange of data without people having to worry about what character set must go with it

- A single code-point (number) for each character

Mostly complete for Western languages

Unicode …

Incorporates basic ASCII

Follows international standard ISO 10646

Has about 60000 characters now

Mostly two-byte

It is independent of language

- Language using the same symbols use the same code points in Unicode

Unicode issues

Evolving as languages evolve

Does not address sorting, font and layout

Contains some ‘private use areas’

Has some idiosyncrasies:

- E.g. identical glyphs with multiple code points
- Some characters can be encoded as a single character or as two
  - —E.g. Ä or A + ¨

Unicode code set vs. format

Each character has a number, but there is more than one way to encode the numbers in data!

Fixed-width

- UCS-4: All characters take 4 bytes.
- Unused bytes set to zero
  - —e.g. US-ASCII up to 128
- Causes considerable expansion of English text

Unicode code set vs. format

Variable-width

- Uses from 1 to 6 bytes
- US ASCII encoded on 1 byte
- Other single-byte characters on 2 bytes
- Most Asian characters on 3 bytes

Unicode fonts

20-30 MB!

Some fonts build in ‘intelligence’

- E.g. how to render text

Some web resources on Internationalization and Localizatioon

W3C: http://www.w3.org/International/

Software Globalization: http://www.wilsonmar.com/i18n.htm

Language Automation: http://www.lai.com/l10ninfo.html

Do a web search and you will find tons of other resources