

# All About Internationalization

## ■ Overview

If you plan to offer your product in other languages, “internationalization” is a critical element in your overall development process.

Internationalization is the first step of a two-step process. It consists in generalizing your product to be as language-independent as required. The second step - localization - consists in adapting the product to meet the needs of different languages and cultures. Internationalization reduces cost and time-to-market by making localization easier and avoiding work duplication.

Unfortunately, the complexity of internationalization is often underestimated, resulting in missed deadlines and cost overruns. This complexity arises primarily from the large number of issues involved and from their inherent pervasiveness: language issues can potentially affect every system, every component, every third-party tool, every line of code, every document, every help file, every test script, and every business process involved in releasing your product to the world.

Our workshop is designed to deliver maximum value in a minimum amount of time. The class is hosted at the location of your choice, either at your site or at a remote conference facility to minimize distractions.

## ■ Target Audience

- The first day is intended for all audiences: programmers, architects, testers, buildmasters, managers, product managers, R&D directors, executives, customer support, sales, etc. Technical concepts are presented in a non-technical manner using visually intuitive and compelling graphics.
- The remaining time is targeted more for technical staff, but much of it remains understandable by all, again due to the visually intuitive mode of presentation.

## ■ Benefits

This workshop will prepare you for all aspects of an internationalization project. You will know the issues, you will know the pitfalls and you will know the solutions. The workshop will provide you with a clear understanding of industry best practices, how to apply them and what their benefits are.

All attending stakeholders within your organization will share a *common terminology* which is the most critical ingredient to enable clarity of design, good communication and true understanding.

## ■ Duration

The agenda described below is for a two day session. The first day (chapters 1 to 7) can also be used as an independent introductory module for all stakeholders.

## ■ Pre-requisites

None.

## ■ Agenda

DAY 1 - for all audiences: executives, product managers, developers, testers, support...
--

### 1. The Problem Domain: What Are We Up Against?

The characteristics of different languages and writing systems are presented visually. Particular attention is paid to the requirements of Asian and Arabic languages. Differences in cultural conventions are also highlighted. This chapter tells us WHAT issues programmers, webmasters and managers should be concerned with (and often are not aware of).

- The globalization dimension
- Languages and scripts
- Asian & Middle-Eastern languages
- Visual presentation of linguistic & cultural differences that impact development
- Cultural formats: date, time, calendars, currency, etc.

### 2. Key Globalization Concepts

Internationalization and Localization. Translation and what translators need. Levels of internationalization complexity. This chapter tells us WHY we should be concerned with the issues presented above, i.e. WHY we should prepare for translation.

- Globalization, Internationalization, Localization, Translation
- Internationalization complexity: dimensions and levels
- Ensuring translation quality
- Maintaining a global product: localizing the next version!
- Optimizing localization: glossaries & translation memories

### 3. Graphical Model of Internationalization

An animated graphical model that illustrates the three basic tasks of internationalization: resource externalization, cultural generalization and preparation of the localization kit. Each of the three tasks

is explored in more detail. This chapter shows us HOW, in general terms applicable to both software and Web sites, to prepare for translation

- Graphical model of core internationalization tasks
- Resource externalization
- Cultural generalization
- Localization kit generation and strategy

#### 4. Internationalization Project Management

From the business case to customer support, through requirements, architecture & analysis, development and testing. Avoiding the pitfalls; managing risk.

- Markets: countries, languages & audiences
- Business case and ROI
- Risk management: the “baseline and compare” strategy
- Evaluation & planning; roles and responsibilities
- Simultaneous Shipment (“sim-ship”) & Parallel development
- Common myths and pitfalls

#### 5. Brief Overview of Development Issues

Basic development concepts for all audiences; more details and depth during day 2 of the workshop.

- Character sets
- Locales
- Resources
- Sorting/searching
- Databases

#### 6. Do's and Don'ts

Numerous practical examples from real projects, visually illustrated.

- Dimensions of evil: concatenation, alignment, abbreviation
- Implicit and explicit concatenation
- Message formatting
- Indexes and sorted lists

## 7. Testing internationalization

A brief and visual overview of internationalization testing issues for all (programmers, managers, etc.)

- Internationalization testing and localization testing.
- Staffing and techniques.
- Regression tests and pseudo-translation.
- Platform coverage and bug workflow.

DAY 2 - for technical staff: architects, programmers, web developers, test script writers...

## 8. Character Sets

- Brief history of character sets
- Single-byte character sets
- Multi-byte Asian character sets on Windows and Unix
- Unicode 4.1: a 21-bit character set

## 9. Unicode: Character Set & Standard

- The 10 Unicode design principles
- The Unicode standard
- Unicode Technical Reports
- UCD: Unicode Character Database
- Unicode resources: tools, books, etc.

## 10. Unicode Representation: Choosing the Proper Form

- Standard encodings: UTF-8, UTF-16, UTF-32
- Encoding size comparisons
- CESU-8: the database patch!
- Serialization & the Byte-Order Mark (BOM)
- Compression schemes: SCSU & BOCU-1
- Data stability: Unicode normalization
- The four Unicode normalization forms

## 11. Unicode Implementation

- The Internationalization Reference Model
- Transcoding and its complexity for bidirectional languages
- Unicode tools and libraries
- Case Handling & Case Mapping
- Text Boundaries: Grapheme Clusters, Words, Lines, Sentences
- Collation-Based Processes, Multilingual Collation Standards

## 12. Locales & Resources

- What is a locale?
- Locale vs. language
- Locale vs. country
- Unicode language and country resources
- CLDR
- The new locale standard
- Resources
- Organizing & naming resource files
- Organizing & naming resources
- Automating resource externalization

## 13. Input and Output

- Input methods: types and frameworks
- Input method programmatic usage
- Fonts, font types, Unicode fonts, font linking
- Text rendering issues and the Unicode bidi algorithm

## ■ Handouts

Each attendee will receive a 200+ page booklet, one slide per page, with ample room to take notes, complete with table of contents. The booklet is designed to serve as a practical useable reference document for on-going use during an internationalization project.

## ■ About our Instructor – Pierre Cadieux

Pierre Cadieux is a veteran with over 20 years experience in internationalization of software, Web sites and embedded systems. He teaches internationalization at the University of Montreal. Pierre has been technology editor for the LISA newsletter, VP Technology at ALIS and director of technology at Bowne Global Solutions.

At ALIS, Pierre pioneered the transparent handling of Arabic and Hebrew languages and created the core bi-directional technology licensed by Microsoft.

As Director of Localization Technology at Bowne Global Solutions, he carried out research and analysis on multilingual Web sites and published the first generic model of Globalization Management Systems.

Additionally, Pierre holds a B. Sc. and M. Sc. in Computer Science.