All About Arabic *for Programmers*

Overview

The Arabic script is *the* most difficult to deal with in software.

This workshop teaches the key aspects of Arabization i.e. the process of adding Arabic support to software systems. It starts with a brief overview of the Arabic culture and follows with the Arabic writing system and its contextual rules, complex ligatures, harakat (vowels), numbers and bi-directional display, etc.

The workshop then covers Arabic data representation in depth: characters sets, encodings, contextual forms, Unicode Arabic blocks and the specific encoding issues for all character classes: letters, digits, neutrals, mirrors, lam-alef, harakat, etc.

With this foundation in place, it moves on to the Unicode Bidirectional Algorithm (UBA) describing basic concepts, rules, character classes, testing tools, development issues, etc. A chapter of common UBA problems and fixes is also presented.

Two more chapters focus on Input and Output. Input concerns the complexities of Arabic data entry and editing, including strange UBA side-effects and how to mitigate them. Output is about modern UI design issues including common controls, font selection, justification, etc.

The remaining chapters review Arabic/bidi support on major platforms: one chapter each for the Web, Android and IOS.

Target Audience

This course is intended for software/app developers, web developers, testers and team leaders, or anyone involved in Arabic support that has some technical background.

Benefits

After taking this workshop you will be ready to start arabizing your product. You will know the main issues and common pitfalls and you will know the solutions. You will know what requirements to consider, what changes your software or Web site or app requires, and how to implement those changes.

Duration

The agenda described below is for a 2-day session.

Pre-requisites

This workshop presumes that attendees have already taken the "All About Internationalization" workshop, either onsite or in the self-paced eLearning format.

Agenda

1. Arabic Culture

A brief overview of the Arabic culture including the various bilingual Arabic markets.

- History of the Arabic script, calligraphy, religion
- Arabic in today's world

2. Arabic Writing System

Presentation of the Arabic writing system: letters and numbers.

- Basic Arabic alphabet (abjad), Persian vs. Arabic, Romanization
- Contextual shapes, kashida (tatweel), harakat, lam-alef and other ligatures
- Numeric display, digit shapes, decimal and thousands separators, mathematics

3. Arabic Data Representation

An in-depth look at Arabic data representation in shape encodings, legacy encodings and Unicode.

- Logical vs. visual order, contextual forms vs. characters
- Legacy encodings (ASMO, ISO, Windows) & the Unicode Arabic blocks
- Representation and transcoding for: letters, harakat, numbers, neutrals, mirrors…

4. The Unicode Bidirectional Algorithm (UBA)

An in-depth presentation of the Unicode Bidirectional Algorithm including all Unicode bidirectional character classes and bidirectional controls.

- Basic concepts: Base direction, language insertions, directional runs
- Neutral character rules and mirror characters
- Numeric classes, rules, examples
- Bidi controls: RLM, LRM, RLE, LRE, RLO, LRO, PDF.
- New Unicode 6.3 directional isolates: LRI, RLI, FSI, PDI.

5. Common UBA Problems & Solutions

A taxonomy of UBA problems that highlights all the common ones.

- Problems with language insertions: hanging neutrals, insertion reversal...
- A real problem with parentheses
- Problems with numeric insertions, phone numbers, MAC adresses
- A real problem with Google Maps
- Bidi problem & solution patterns.

6. Arabic Data Entry & Editing

The numerous quirks of bidi editing.

- Arabic keyboard layouts and usage
- Language and keyboard control under Windows
- Base Direction
- Bidi text entry and editing
- Bidi cursor and selection
- Bidi surprises in copy & paste!

7. Arabic Data Display

Practical tips on designing an Arabic User Interface.

- Arabic fonts: harakat support, ligatures, font pairing, font attributes
- Images: translatable, directional and cultural
- Mirroring of UI widgets: navigation, undo/redo, progress bars, sliders…
- Mirroring of UI layouts: web pages, graphs, bidi forms...
- Line wrapping and justification

8. Arabic on the Web

Theory and pratice of Arabic handling on the Web.

- Core concept: encoding, language, direction
- Arabic HTML
- Arabic CSS
- A practical case: Bidi-Ing the Test Bench
- Common Web Problems and Solutions

9. Arabic on Android

Theory and pratice of Arabic handling on Android.

- Getting Started
- Locales & Pseudolocales
- Bidi UI specifics
- Arabic & RTL Resources
- UBA Support
- Fixing Bidi Problems

10. Arabic on iOS

Theory and pratice of Arabic handling on iOS.

- Getting Started
- Locales & Pseudolocales
- Bidi UI specifics
- Arabic & RTL Resources
- UBA Support
- Fixing Bidi Problems

Handouts

Each attendee will receive a 600+ page booklet, with ample room for notes, complete with table of contents and glossary. The booklet is designed to serve as a practical easy-to-use reference "book" for regular use during an internationalization project.

About our Instructor – Pierre Cadieux

Pierre Cadieux is a veteran with over 35 years' experience in internationalization of software, Web sites and mobile devices. He has taught internationalization at the Université de Montréal. 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.