
Unicode in Mobile Phones



Adil Allawi

Technical Director

adil@diwan.com

Unicode in Mobile Phones

- **Complex text in small devices**
 - Memory is very limited
 - Processing speed is low yet time is critical
 - Screen is small
- **Unicode is common on mobile phones**
 - But, the problems come with displaying complex languages.
 - Most devices are first designed for simple left to right text
 - Later the software is converted to handle multilingual text



1

Why Diwan and Mobiles?

■ Started with Arabic Electronic Publishing Revolution in 1984



- Worked with Apple Computer Inc. to define new fonts and Arabic OS interface
- First Arabic Desktop Publishing
- First Fully Electronic Newspaper

■ New Development in Mobile Computing



- Arabic WAP publishing system as Images
- Specially developed Arabic typeface for small devices
- Working with Siemens, Openwave and others to define complex text support in mobile phones.

1

It's Like Deja-Vu all over again!



- 1984 - Macintosh
 - an 8Mhz processor
 - a 512x342 b/w screen
 - 128Mb RAM
 - 800Kb OS, and Applications.

- 2002 Mobile phones
 - Very similar!



- But there are differences
 - screen sizes can be much smaller
 - Input is more difficult and good presentation of text more critical
 - Feedback needs to be instantaneous

1

Presentation of Text

- Can be limited to a minimal set of characters
 - Just the characters that are needed for the input method (e.g. T9)
- Need to present as much text as possible in a small space
- Can compromise on language
 - e.g. Arabic does not need accents
- Font design is very important

1

BiDi Text Support

- Unicode bi-directional algorithm
 - Can do without support for directional characters
- Basic set of characters needed
 - No need to present accents
 - Only need to support essential ligatures
- Specially designed Arabic font

أنواع الخطوط العربية: الإسماعيلي، والمكي، والمدني، والأندلسي، والشامي، والعراقي، والعباسي، والمشعب، والريحاني، والمجرد، والمصري، فهذه هي الخطوط العربية التي كان منها ما هو مستعمل قديماً ومنها قريبة الحدوث، أما هذه الطرائق المستنبطة فهي مروية عن الصحابة حتى اتصلت بابن مقلة وياقوت وغيرهم. وهم تفننوا فيها حسب اجتهادهم.

أنواع الخطوط العربية: الإسماعيلي، والمكي، والمدني، والأندلسي، والشامي، والعراقي، والعباسي، والمشعب، والريحاني، والمجرد، والمصري، فهذه هي الخطوط العربية التي كان منها ما هو مستعمل قديماً ومنها قريبة الحدوث، أما هذه الطرائق المستنبطة فهي مروية عن الصحابة حتى اتصلت بابن مقلة وياقوت وغيرهم. وهم تفننوا فيها حسب اجتهادهم.

1

BiDi Selection

■ Visual vs Logical Highlighting

“It’s Mine” قال النمر
 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

“It’s Mine” قال النمر

Logical Highlighting - characters 6 to 13 are selected

“It’s Mine” قال النمر

Visual Highlighting - characters 6 to 10 and characters 15 to 19 are selected

1

Thai Language Support

■ Main Features of Thai rendering

	1	TOP
	2	ABOVE
Ag๓๓๓๓	3	BASE
	4	BELOW

Thai language requires combining characters. A combined character may be displayed on more than one level.

๓
พพ

Combined characters may overlap the following character.

พำ๓

Correct display of illegal character combinations

1

Thai Language Support

- Text engine correctly places Thai accents in their optimal location allowing clearer display of Thai text.
- Thai fonts can be drawn legibly to a minimum of 10 pixels - ideally 13 pixels
- Thai numbers may be displayed instead of European numbers or as an alternative.

Thai Input Method

- Thai input works in two modes
 - **Passthrough**: allows entry of all letters without checking.
 - **Strict**: will prevent the entry of illegal combinations.
- Cursor movement can jump over groups of combined Thai or can be drawn between combined characters
- Compatible with Thai T9 Input

Thai Word Breaking

- Thai language is commonly written without word breaks.
- Guessing of correct word breaks is non-trivial.
- Current implementations need a dictionary of 92K but this needs to be smaller.

1

Fitting it all in

- Complex text possible in a small space
 - Araic and Thai fonts about 12Kb
 - Bidi Algorithm 8Kb
 - Arabic and Thai shaping 8Kb
- Maximum time added is 10%

Designing Software

- Display of complex text
 - Display glyph may be different from the input characters
 - Characters may be reordered
 - Ligatures may replace several characters
- Software should account for this:
 - Measure routine should be passed whole paragraph - not just characters to be measured
 - Separate measuring for line breaks from measuring the cursor
 - Don't assume words are divided by certain characters
 - Use 4 separate cursor routines:
 - Left, Right, Previous, Next
- Layer Drawing code
 - Functions should deal with text at the character level another at the line level etc.

1

Mobile Java Applications

- Current implementation needs support from phone to handle complex text.
- This can give different results across devices
- Some work should be done by the VM.
 - e.g. allow application to specify the Bidi dominance

BiDi WAP

■ Three levels of BiDi WAP support:

1 🏠 WHATS ON
 2 💬 CHAT
 SOME TEXT:
 MARY HAD a little lamb ITS
 FLEECE WAS WHITE AS
 SNOW. NUMBERS: 012345.

This is the page if drawn before processing

1 🏠 NO STAHW
 2 💬 TAHC
 :TXET EMOS
 STI a little lamb DAH YRAM
 SA ETIHW SAW ECEELF
 .012345 :SREBMUN .WONS

This is the page as displayed with BiDi text enabled but no ordering

NO STAHW 🏠 1
 TAHC 💬 2
 :TXET EMOS
 STI a little lamb DAH YRAM
 SA ETIHW SAW ECEELF
 .012345 :SREBMUN .WONS

This is the page as displayed with full BiDi ordering



BiDi WAP

- General Design notes to stay compatible with current and future WAP sites
 - Keep plain text plain
 - No pictures in line
 - No formatting of part of the text (e.g. emboldening a single word)
 - Avoid Tables for right to left content
 - Use Unicode encoding

1

Thank You!



Adil Allawi

Technical Director

adil@diwan.com