

Efficient Algorithms for the integration of Arabic Language in Mobile Phone

A. ABDELHADI, L. H. MOUSS and O. KADRI

Abstract—This work appears in man-machine interface. Our goal is to study the integration of Arabic language in Mobile Phone, in order to achieve a man-machine interface Arabic. The correct display of Arabic character is essential in a MMI. Since Arabic characters change their forms according to their position in a word, then it is necessary to make a contextual analysis on every word, to find the correct form of each character. The transformation of two or more characters in one form, demand special treatment, as in the case of Arabic ligature LAM-ALEF. The Arabic language has a different direction of writing in relation to other languages embedded in mobile phone, which requires finding an algorithm that provides a bidirectional display of SMS messages. These messages may contain characters from different direction, from right to left, left to right or characters that have no direction. It allows you to make the message understandable.

Index Terms—characters, contextual analysis, Glyph, Ligature, Bidirectional Algorithm, logical order, visual order.

I. INTRODUCTION

The Arabic language has several features that require special treatment to be implemented in a program, integrated into a PC or a mobile phone. First, the direction of writing is from right to left. In addition, Arabic characters change their forms according to their position in a word. There is also the possibility those two characters in one form. This article discusses the major impacts to integrate an Arabic MMI (Main Machine Interface) in mobile phone.

The graphical interface of a mobile phone is composed of a set of dialogs called screens. A screen contains one or more strings called prompts. In the case of Arabic MMI, The prompts consist of a string of Arabic characters.

The impact of the Arabic language on mobile phone is mainly at the level of the display of Arabic characters. Because the Arabic message written from right to left while the character code is stored from left to right. A contextual analysis on the Arabic prompts is required to be displayed correctly because according to the position of the Arabic character in a word, it may have a different form or, at the beginning, middle, at the end or isolated. The transformation of two characters in one form is possible in the Arabic language.

Manuscript received December 23, 2010; revised April 12, 2011.

A.ABDELHADI is with the Laboratory automation and production, Department of Industrial Engineering, University of Batna, 1, Rue Chahid Boukhrouf 05000, Batna, Algeria (phone: 21333803396; fax: 21333803396; e-mail: Abdelhadi.adel@gmail.com).

L.H.Mouss is with the Laboratory automation and production, Department of Industrial Engineering, University of Batna, 1, Rue Chahid Boukhrouf 05000, Batna, Algeria (e-mail: hayet_mouss@yahoo.fr).

O.Kadri is with the Laboratory automation and production, Department of Industrial Engineering, University of Batna, 1, Rue Chahid Boukhrouf 05000, Batna, Algeria (e-mail: ouahabk@yahoo.fr).

However, this transformation requires a special treatment to ensure the correct display of this form.

The MMI of mobile phone offer menus, screens, but the possibility of publishing and displaying short messages named: SMS (Short Message Service). A text message consists of a chain of character belonged to a single language or several. The editing and display a bilingual message SMS needs a bidirectional display algorithm.

II. ARABIC IN COMPUTING

The Arabization of computer systems Single or Multi-User affects the two fundamental aspects of the material on the one hand and programs, including programming languages, utilities and operating systems on the other.

Arabization is geared towards attaining the following objectives:

- Making use of the Arabic language in computing,
- Making use of the Arabic language in computing as simple and efficient as the Romance languages, Designing applications independent of the language of dialogue,
- Manage data in multiple languages simultaneously, Design applications fully independent of the language of dialogue chosen by the user and integrating the different languages simultaneously. [BEN 03].

A. Computer coding of the Arabic alphabet

The standard ASCII coding has shown its short comings, since it works on 8 bits, i.e. it allows only 128 coding positions. Currently there are over a million characters in the world who need coding to meet the increasing demands of industrial language non-English-speaking world, and for other languages such as Arabic and its rules links to appear on a computer screen.

This shortfall has forced computer manufacturers to Create another coding standard, which can handle this huge number of character, but compatible with existing standards is the Unicode or UCS. This standard was created by a group of computers in 1989. It allows you to set the encoding for the majority of characters used in languages of the world. It is a game of encoding two or more bytes, and to ensure compatibility with the standard ASCII, the first 256 characters are reserved for ISO-Latin-1 [Fan 99].

B. Arabization of mobile phone

The base of the Arabization of a system is the adaptation of all aspects of the operating system to allow the user to manage the Arabic text or input or the output. Specifications Arabization advanced in the field of mobile phone may include many other features such as predictive (as a working tool to limit the number keypad of the phone).

The basis of the request Arabization to answer two main

questions includes:

- **At the input:** how the user responds with the system to enter an Arabic text?
- **At the output:** how the system will display the Arabic data that have been input by the user? And some other problems such as:
- **Alignment of Text:** accept and manage a left- aligned text (Latin) or just right-aligned (Arabic).
- **Cursor management:** who is used to treat the cursor position, either before or after a character in bidirectional text?

The requirements for handling all these specifications are:

- Bidirectional treatment,
- Change glyphs,
- Management of vowels (CHACL),
- Arab font with size supported by the system.

Figure 1 shows the system of mobile phone:

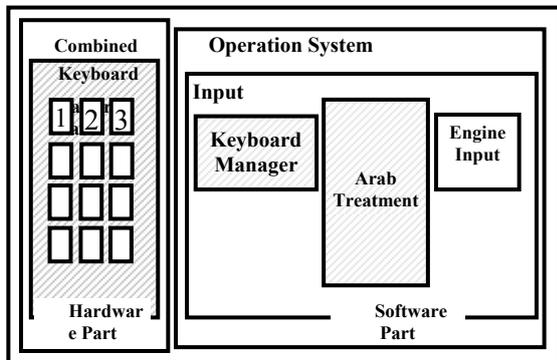


Fig. 1 Mobile phone system

Figure 2 shows all the modules to build a mobile phone system and all their connections. The typical scenario, described below, where all system modules are involved, shows all the steps of printing an Arabic message.

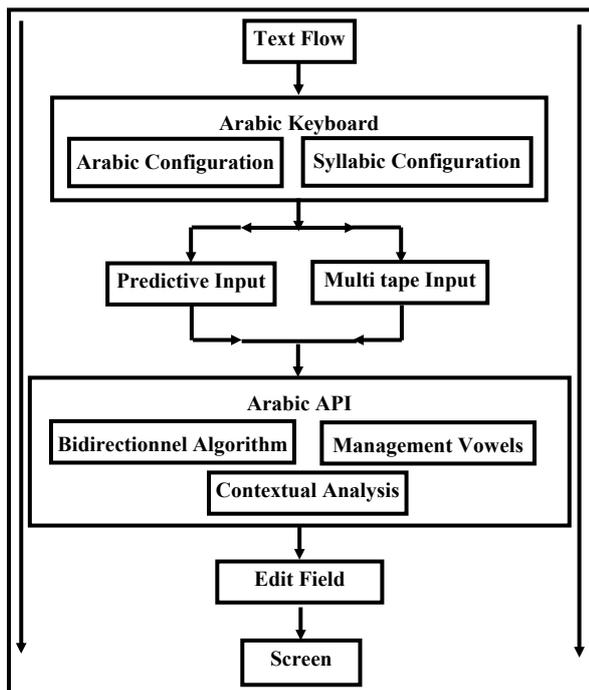


Fig. 2 Interior Description of mobile phone system

III. CONTEXTUAL ANALYSIS ALGORITHM

This section explainthe different stepsof development of the contextual analysis algorithm (Glyph SHAPING). This algorithm allows changing the glyph of each Arabic character based on its position in a word for it to appear correctly when viewed on the screen of the mobile phone.

Our algorithm is used as part of a more general algorithm for bidirectional display text in SMS message.

A. Principe of the algorithm

Contextual analysis is required to write several languages in order to present their characters with correct form. For the Arabic language, to choose the appropriate form of presentation of a character, it is necessary to take into account the form of their characters if they are neighbors.

The Glyph SHAPING algorithm is based on four paintings depicting different types of glyphs that can turn into Arabic.

B. Automate of Contextual analysis algorithm

In Arabic, the letters of a series gradually change their forms with the use of contextual analyzer. Indeed, only the last two letters of this sequence change. We use finite state machines to implement our algorithm. The states are forms of the last two letters, and transitions are labeled by the type of the previous character. The previous shows that before the last character read by the analyzer. Only two actions are possible: "the nature of the class (L)" which means that the previous character is the character table of the class L, and the character of the class (R), which means the opposite.

Figure 3 shows the graph associated with our automatic contextual analysis:

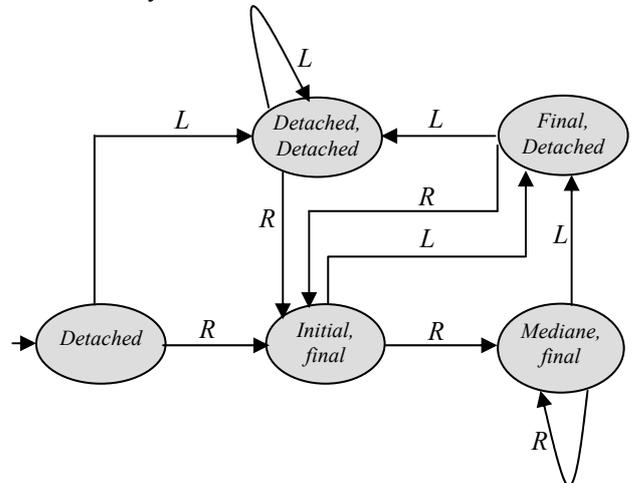


Fig. 3 Graph of automated contextual analysis

The contextual analyzer receives a context with the detached form. So the initial state consists only of the first form letter. There is a transition from one state to another after playing a new character. The robot stops when the series ends. Thus, all states can be seen as final statements.

The following table shows an example of using this finite state automaton:

TABLE 1 EXAMPLE OF CONTEXTUAL ANALYSIS

Analyze the word: نظام
Read the character = ن)
Analyze = ن) ("detached" stat= initial stat)
Read the character = ظ

The previous character () is a class
R The transition "R" is selected
 Analyze = نطق (stat "initial, final")
 Read the character = ل
 The previous character (ط) is a class
R The transition "R" is selected
 Analyze = نطق (stat "median, final")
 Read the character = م
 The previous character (ل) is a class
L The transition "L" is selected
 Analyze = نظام (stat "final, detached")

IV. BIDIRECTIONAL DISPLAY ALGORITHM

This section explains the different steps of the Bidirectional Display Algorithm (BIDI MOBILE). This algorithm allows for a string to display on the screen of the mobile phone, which is written in two different directions: from right to left (like Arabic), from left to right (like the Romance languages) and that can also solve the problem of characters that have no direction (neutral). This algorithm is used in the integration of Arabic in mobile phone. It is used to prepare the string constant (prompts) and a MMI also provides publishing and display of SMS messages.

A. Principe of the algorithm

Before starting the algorithm display bidirectional (BIDI MOBILE), the first thing to do is treat the route that lies between the Arabic characters, using the contextual analysis algorithm (Glyph SHAPING). This algorithm allows replacing the Arabic characters by their glyph correct.

There is another task independent of the bidirectional algorithm, and is the operation division of the treaty text online. The proposed solution is based on the property of directionality of each character.

This solution is divided into three basic parts:

- Defining the display directionality: this part is to define directional type for each character and specify the display direction.
- Display Direction right to left: if the type of first character of the input string has R type, then the direction of the entire output from right to left.
- Display Direction left to right: if directional type of the first character of the string input has L type, then the output string is left to the right.

B. Defining the display directionality

1) Definition of directional type of characters

The definition of directional type of each character of the string input is essential for the correct display of the message. The proposed coding is as follows:

TABLE 2 DEFINITION OF DIRECTIONAL TYPE FOR EACH CHARACTER

R	Arabic and Hebrew characters	0
L	Most of alphabetic characters	1
AN/EN	Number	2
N	Neutral Character	3

2) Definition of the database management display

In this section, we define a basis of display direction for the input string. This treatment is based on the direction type of first character encountered strong direction and allows

distinguishing these two cases:

- Display direction right to left.
- Display direction left to right.

C. Display direction right to left

First, the direction of the string input character is from right to left. Our solution is based on a finding the whole chain of successive characters of R type and the characters of N type.

In this case, there are several rules of resolutions neutral characters to be used to give a direction to each character of N type.

In the event there is an ambiguity in defining the direction of a character of N type, as shown in the following two rules:

TABLE 3 DIRECTION OF THE CHARACTER OF N TYPE

R	N	eor	→	R	i	eor
soR	N	L	→	soR	i	L

The character in question (i) takes in this case the warp direction of current input character.

Then, we must find the correct way to display these characters. The solution is to reverse completely the string found, and bring it to the end of the output string.

In all the examples will be presented:

- XX: uppercase characters represent Arabic characters.
- xx: the lowercase letters represent the Latin alphabet

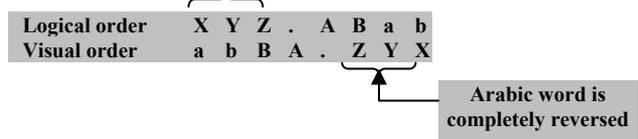


Fig. 4 Conversely every Arabic word

The search for the string to be displayed from right to left may contain characters of R type and N type.

For an N with type character belongs to the string R, it must verify certain conditions, so they are not between:

- Two characters with L type,
- Two numbers ,

One of the two terminals is a character with L type and the other is simply a number or vice versa.

The treatment applied to a sequence of numbers or characters of L is the same. First determine its location, i.e. find the start and end in the chain. Second, the reorganization of the chain is done as follows:

The chain takes the same order as in the string of entry, but with a lag, depending on the presence or absence of a string R in the beginning.

TABLE 4 DISPLAY DIRECTION RIGHT TO LEFT

Example1	Logical order	X	Y	a	b	c	1	2		
	Visual order			a	b	c	1	2	Y	X
Example2	Logical order	A	B	1	2	3	4	5	6	
	Visual order			1	2	3	4	5	6	B

According to the previous examples, we find that the words of direction R are completely reversed by words against the direction L and the numbers remain the same with a lag.

D. Display direction left to right

This section allows you to search for a string of L characters, numbers, or neutral.

The neutral must verify certain conditions so that they are between:

- Two characters with L type,
- Two numbers,
- One of the two terminals is characters with L type terminal and the other is simply a number or vice versa.

TABLE 5: DISPLAY DIRECTION LEFT TO RIGHT

Case 1	Logical order	a	b	c	d	X	Y		
	Visual order	a	b	c	d	Y	X		
Case 2	Logical order	a	b	C	D	a	3	,	4
	Visual order	a	b	D	C	1	3	,	4
Case 3	Logical order	a	b	1	c	d	Y	Z	
	Visual order	a	b	1	c	d	Z	Y	

The process of reorganization of the string shall be as follows: A sequence of characters L and remains neutral in

the same order without lag. As against the strings R are reversed and set next to the string L.

V. RESULTS OBTAINS

Several tables show a comparison between the results obtained by BIDI MOBILE and the results provided by the reference to Unicode.

We can see three possible:

- Display direction right to left.
- Display direction left to right.
- Display direction neutral.

XX: uppercase characters represent characters in direction R.

xx: lowercase characters represent characters in the direction L. N: the rest is neutral characters

TABLE 6: RESULTS OBTAINS WITH DISPLAY DIRECTION RIGHT TO LEFT

Test	source string	Unicode Reference	Results BIDI MOBILE
1	CAR IS the car IN ENGLISH	HSILGNE NI the car SI RAC	HSILGNE NI the car SI RAC
2	NUMBER IS 123456	123456 SI REBMUN	123456 SI REBMUN
3	TEST 23 ONCE abc	abc ECNO 23 TSET	abc ECNO 23 TSET
4	TEST ABC 123456	123456 CBA TSET	123456 CBA TSET
5	SOLVE 1*5 1-5 1/5 1+5	1+5 1/5 1-5 5*1 EVLOS	5+1 5/1 5-1 5*1 EVLOS
6	HE SAID "it is 123, 456, ok"	"ok ,456 ,123 it is" DIAS EH	"ok ,456 ,it is 123" DAIS EH
7	HE SAID "it is a car!" AND RAN	NAR DNA "lit is a car" DIAS EH	NAR DNA "lit is a car" DIAS EH
8	HE SAID "it is a car!x" AND RAN	NAR DNA "it is a car!x" DIAS EH	NAR DNA "it is a car!x" DIAS EH
9	THE RANGE IS 2.5..5	5..2.5 SI EGNAR EHT	5..2.5 SI EGNAR EHT
10	TEST ~~~23%% ONCE abc	abc ECNO 23%% TSET	abc ECNO %%23 TSET

TABLE 7: RESULTS OBTAINS WITH DISPLAY DIRECTION LEFT TO RIGHT

Test	source string	Unicode Reference	Results BIDI MOBILE
1	car is THE CAR in Arabic	car is RAC EHT in arabic	car is RAC EHT in arabic
2	number is 1234567	number is 1234567	number is 1234567
3	he said AB "IT IS OK 123"	he said "123 KO SI TI" AB	he said OK SI TI" BA 123 "
4	he said "IT IS 123, 456, OK"	he said "KO ,456 ,123 SI TI"	he said "SI TI 123, 456, KO"
5	he said "IT IS (123, 456), OK"	he said "KO ,(456 ,123) SI TI"	He said "OK (456 ,123) IT IS"
6	he said "IT IS A CAR!X"	he said "X!RAC A SI TI"	he said "X!RAC A SI TI"
7	he said "IT IS A CAR!"	he said "RAC A SI TI"	he said "RAC A SI TI!"
8	abc (TEST)	abc (TSET)	abc)TSET(
9	he said "A SI TI bmw KO ,500."	he said "A SI TI bmw KO ,500."	he said TI SI A " bmw KO ,500."
10	a/1	a/1	a/1

VI. CONCLUSION

The aim of our work was to study the impact of integration of the Arabic language in mobile phone, to arrive to conduct a man-machine interface within the reach of the Arabic world. In this article we describe mainly the implementation of several solutions to solve the problems posed by the characteristics of the Arabic language.

Our work presented in this article on the explanation of the steps followed for the development of two algorithms, the contextual analysis algorithm (Glyph Shaping) and bidirectional display algorithm (Bidi Mobile).

The contextual analysis algorithm designed primarily to address the route that lies between the Arabic characters and define their correct form, whatever the type of characters

neighbors, Arab or other. Note that each character can take several forms. This algorithm can also solve the problem of Arabic ligature LAM- ALEF, and that during the analysis of all characters. This algorithm is used by another algorithm, is that of the bi- directional display.

The bidirectional display algorithm is an algorithm to treat the complications of writing the Arabic language. This solution is essentially the field of mobile phone, and more particularly, to find a solution for the correct display of SMS messages.

We believe that the objectives of this work have been achieved. Moreover, the solutions proposed to address the characteristics required by the Arabic language are effective, easy and within reach of users who use mobile phone integrating these solutions.

However, these results do not exclude that some additions may be made to our work. This concerns in particular the following points:

- The treatment of vowels associated with Arabic characters which are generally used in religious texts and materials. However, the use of these vowels in the SMS messages may remove the ambiguity between certain words.
- Improving the bidirectional display algorithm to deal with the complicated bilingual SMS messages, which contain neutral characters, and have the mirror property.
- The processing of SMS messages representing arithmetic operations written in Arabic, from right to left.
- Consider the use of dictionaries of Arabic words, and this level of publishing SMS, using the predictive character.

- [5] M.Fanton, "TEX : les limites du multilinguisme", centre d'étude et de recherche en traitement automatique de l'INALCO associée au CNRS, paris Septembre 1999.
- [6] D.Grobgeld, "A Free Implementation of the Unicode Bidi Algorithm", 2003
- [7] E. Hart, "The Unicode Character-Glyph Model: What you Need to Know about Processing and Rendering Multilingual Text", 15th International Unicode Conference, San Jose, California, August30-September 2, 2004.
- [8] A. Jacques, "Caractères, codage et normalisation – de Chappe à Unicode", 2004.
- [9] M.Leisher, "The UC Data Unicode Character Properties and Bidi Algorithm Package." July 17, 2004.
- [10] A.Patrick, "Introduction à Unicode et à l'ISO 10646", 2002.

REFERENCES

- [1] S.Atkin and R., "Stansifer Implementations of Bidirectional Reordering Algorithms", 18th International Unicode Conference Hong Kong, April 2004.
- [2] M.Benhenda, "vers une normalisation des pratiques de communication dans le contexte d'un multilinguisme intégral (arabe latin) ", Institut Supérieur de Documentation, Tunisie, 2003.
- [3] M.Davis, The Bidirectional Algorithm, Unicode Standard Annexe9, mars 2005.
- [4] A.Dean. "Optimized Implementations of Bi- directional Te Layout and Bi-directional Caret Movement." Thirteenth International Unicode Conference, September 1999.

Abdelhadi Adel received the magister degree in 2004 from the Department of Computer Science, University of Batna, Algeria. He is currently an assistant professor at the University Center of Khenchela, Algeria. He is currently a Doctoral student in the Department of Industrial Engineering, University of Batna, Algeria. His research interests include preventive maintenance, multi-agent system, classification, diagnosis and artificial immune system.

Mouss Leila-Hayet is professor in the Department of Industrial Engineering, University of Batna, Algeria. Here research interests include maintenance, diagnosis, Scheduling and Production.