أبحاث ودراسات بلغات أجنبية

outside the Arab world. Internal and external services shall comprise supplying of required documents, bibliographic and statistical information. One aspect shall include services relating to administration and financial matters.

FARABI shall include three data bases :

- SADIC for documentation and information
- SAIB for statistical data and data processing
- SAFA for administrative and financial matters

The Bank shall also be equiped with appropriate accessories (microfilm and facsimile) to facilitate communication and information exchange with Arab and international data banks, thus featuring an ability of using Arabic and Latin letters. Other units shall be equiped with terminals and printers in order to produce statistical tables, designs and graphic drawings in colored as well as normal forms.

As a training centre, FARABI shall offer continuous in-service-training in the above-mentioned fields for employees in the different departments of Alecso. In a later stage this activity shall be extended to all those working in these fields in the Arab countries, with due consideration to avoiding of training offered by local institutions. In other words the training programmes shall be designed in a way that would complement the individual local programmes.

Finally Alecso considers that the establishment of FARABI is not the end of its ambitions, but a start, at the end of a tunnel, of work which aims at greater and continuous prospects for achievement and whose one of the supporting pillars shall be an overall Arab communication network.

FARABI

FIRST ARAB BANK OF INFORMATION FOR EDUCATION, CULTUTRE AND SCIENCE

ALECSO

Considering the present as an era of communication where all human endeavors are intertwined, it is undoubted that the future shall bear a still greater role to be played by communication in molding newer prospects of human advancement in all areas of knowledge, discoveries and inventions. This fact has driven Alesco to esatblish a data bank and mechanize its activities. As usual, the starting point had been a feasibility study for a project in this connection, which lasted for almost two years. This study benifited, to a large extent, from all possible sources available in the Arab world since 1975, especially the existing limited utilization of informatics. This study bore the fruit of FARABI (First Arab Bank of Information).

Alecso bank of information (FARABI) shall offer various services to the departments and outside offices affiliated to the Organization in addition to istitutions within and

the value of multilingual MINISIS, the Canadian IDRC has already begun the work of supporting alternate character sets within MINISIS.

This support will be done at the intrinsic level within MINISIS, with all input/output operations passing through a new intrinsic dedicated to handling character-set problems. IDRC will be using the HP-2645A terminal with option L98 as the basis for this work, but will write the software in such a way as to make it as independent of the terminal as possible; a configuration processor will allow the user to specify certain characteristics of the terminal in use [17].

Related to the problem of Arabization of hardware and software are problems resulting from the lack of standardization and the tools needed to promote the use of unified standards within Arab library and information centers. Both ALDOC and ALECSO are attempting to promote the use of unified Arabic standards for the bibliographic description of Arabic records.

Conclusion

In search of a standardized approach to the use of Arabic in computerized information interchange, several possible solutions have been put forward for consideration by a number of Arab organizations and institutions. Although the software and hardware which have been proposed do not answer all the problems posed by the Arabic language and culture, there has been a great deal of progress made within the past four to five years. I believe we can expect an important breakthrough in this field in a very short time.

References

1. Hegazi, O.; Ibrahim, Z. I. "A Case Study: Transliteration of Arabic characters in AGRIS." Paper Presented at the First International Information Conference, Cairo, Egypt, December 13-15, 1982.
2. Wellisch, H. H. "Script conversion practice in the world's libraries." *International Library Review*. 8:55; 1976.
3. Kubba, G. A. "The Impact of Computers on Arabic Writing, Character Processing and Teaching." In: *Information Processing 80: Proceedings of IFIP Congress 80*. Amsterdam: North-Holland; 1980: 962.
4. Mosawi, A. "Bilingual problems." *Middle East Computing*. 28; 1982.
5. Library of Congress Cataloging Service. *Bulletin*. 118; 15; Summer 1976.
6. Dewachi, A. "Considerations for the Implementation of an Arabic Code." Paper Presented at the Conference of the Gulf Organization for Industrial Engineering, Doha, May 1981.
7. "Isti'mal al-Shafra al-'Arabiya al-Muwahadda fil-Hasibat al-Iliktruniya" (Use of Unified Arabic Code in Electronic Computers). *al-Mijallah al-'Arabiyah lil-'Ulum (Arab Journal of Science)*. 1(1): 95; 1982.
8. ALECSO. "Final Report on the Use of the Arabic Language in Computers." Paper Presented at Ain Shams University, Cairo, April 13-16, 1977.
9. IBI/COARIN. Minutes of the First Meeting, Rome, June 14-15, 1977.
10. *al-Mijallah al-'Arabiyah lil-'Ulum (Arab Journal of Science)*. 1(1): 95; 1982.
11. *al-Mijallah al-'Arabiyah lil-'Ulum (Arab Journal of Science)*. 1(1): 99; 1982.
12. ISO 646. *7-bit Coded Character Set for Information Processing Interchange*. Geneva: ISO; 1973: 12 pp.
13. ISO 2022. *Code Extension Techniques for Use with the 7-bit Code Character Set*. Geneva: ISO; 1973: 23 pp.
14. ISO 2375. *Data Processing: Procedures for Registration of Escape Sequences*. Geneva: ISO; 1974: 2 pp.
15. Isotta, N.E.C. "The EURAB terminal." *ESA Bulletin*. 16:46-50; 1978.
16. Graham, D. "Saudi-built computer programs in Arabic." *Saudi Business*. 7:22-25; 1983.
17. Valentin, R. *Computer Configuration and Applications for ALDOC*. Paris: UNESCO; 1981: 35.

• Letter quality printer based on the Qume Sprint 5. Printing at 55 characters per second, the 3260 SD handles either Latin or Arabic through interchangeable daisywheel print heads. Support software is on diskette.

Multi-Media Video (MMV)

MMV has produced Arabdata 20, which employs the ASCII Code for Arabic characters using one ASCII Code for each character. It displays English and Arabic simultaneously and performs contextual analysis to determine correct Arabic character shapes. Hard copy output is provided by the Arabprint 160, a dot matrix printer functioning at 160 characters per second.

The Company has developed Arabic versions of the business application software running on the CADO microcomputer and created a bilingual version of the Apple II microcomputer. Called Arapple, the system performs in either Arabic or English, carrying out automatic contextual analysis of Arabic characters.

EURAB

EURAB is an Arabic and Latin Computer terminal patented by ESA (European Space Agency) [15]. This multiple-alphabet terminal allows for:

• writing in Arabic with vowels explicitly displayed,
• the simultaneous display of two pages of 800 characters each and a different alphabet per page,
• up to eight pages stored in the terminal buffer memory,
• independent printing of any number of stored pages while work is in progress at the keyboard.

The LEXAR software transforms the EURAB bilingual terminal into a data entry terminal.

They both form an integral part of the ESA-RECON interactive information retrieval system operated by the ESA Information Retrieval Service.

SK Computer System

A UK Digital Equipment systems house, SK Computers supply the text controller bilingual terminals and printer with DEC systems. SK Computers has made two bilingual installations in Iraq. A range of bilingual software is being developed in conjunction with TTI Computer Systems, a software house experienced in creating Arabic systems.

IBM

IBM has introduced bilingual terminals that are not acceptable to the Arab experts. It has recently entered into an agreement with the Kuwait Institute for Scientific Research to work on problems associated with the use of Arabic in EDP.

ATARI

A recent agreement between al-Alamiya Company in Kuwait and the ATARI Corporation would result in the Arabization of the ATARI Home Computers. The project, when completed, would result in implementing Arabic on various Atari system components.

Saudi Computer Industries

A recent announcement in Saudi newspapers revealed that a group of Saudi professionals intend to market a "homegrown bilingual Arabic computer" to educational institutions that will make computer programming possible for non-English speaking Arabs. The Arab computer, known as al-Farabi, was developed by Saudi scientist Dr. Reda Seraj Al-Thiga, Dean of the Graduate School at the University of Petroleum and Minerals. A group of Saudi professionals and scientists have formed Saudi Computer Industries to build the machine and launch Saudi Arabia's first indigenous computer company. The Arabic computer has built into it the ASCII code. The program can be adapted to any known Arabic code within a short period of time [16].

Unlike IBM and NEC computers which translate English into Arabic and are bilingual, the Arabic computer is programmed in Arabic.

Arabization of Software

Related to the desire of Arab librarians to utilize Arabic/Latin terminals and VDU is their desire to Arabize library specific software. Changes in existing software concern support for the Arabic character set. These requirements include:

(a) The ability to support both Arabic and Latin character records within one database.

(b) The ability to identify fields as being associated with a default character set for prompting and for the data contents which would require additional option in the data definition.

(c) The possibility of embedding character strings of one character set within a field containing text in a second set.

(d) Support for the generation of Arabic keys, including appropriate strippings, key extraction, and key generation options.

(e) The ability to display or print mixed Arabic/Latin text anywhere on a line or page, within the limitations of the software print formatting routines.

The Arab League Documentation Center in Tunis, in cooperation with IDRC and with funding from the UNDP, has begun work on Arabizing the MINISIS Software by Arabizing MINISIS dialogue and messages and translating its documentation. This is the first attempt in the Arab World to Arabize a software that could be used in Arab libraries and information centers. Being aware of

TABLE 4. CODAR-U/FD.

b^4	b^3	b^2	b^1		0	1	2	3	4	5	6	7
0	0	0	0	0			SP	0	@	ذ		
0	0	0	1	1			!	1	ء	ر	ف	
0	0	1	0	2			"	2	آ	ز	ق	
0	0	1	1	3			#	3	أ	س	ك	
0	1	0	0	4			$	4	ؤ	ش	ل	
0	1	0	1	5			%	5	إ	ص	م	
0	1	1	0	6			&	6	ئ	ض	ن	
0	1	1	1	7			'	7	ا	ط	ه	
1	0	0	0	8			(	8	ب	ظ	و	X
1	0	0	1	9			)	9	ة	ع	ى	<<
1	0	1	0	10			*	:	ت	غ	ي	>>
1	0	1	1	11			+	;	ث	ا		
1	1	0	0	12			,	<	ج			
1	1	0	1	13			-	=	ح	[		
1	1	1	0	14			.	>	خ	\		
1	1	1	1	15			/	؟	د	_		

It is advisable for information specialists to determine precisely their requirements in this field and to work with their national computer centers which are actively working on these problems. A case in point is the Centre National de l'Informatique in Tunis which is working on a "definitive" terminal for handling Arabic text, under a grant from IBI. The Canadian IDRC is also looking into this problem in the context of the modifications needed for MINISIS (a minicomputer version of the Integrated Set of Information Systems) and will be using the HP terminal as the basis for the development work to be done. In cooperation with IDRC, the Arab League Documentation Center is presently Arabizing the MINISIS I/O Programs using its Arabic terminals and will be able to use their combined experiences in selecting the definitive Arabic terminals to be used in ALDOC and elsewhere in the Arab world.

Among the manufacturers of bilingual terminals are [4]:

al-Muhandis

This Saudi company introduced its first terminal, the Ascribe, in 1980; its second, Al-Arabi, has recently been released.

Ascribe displays the full-ASCII Latin 96 character set and 112 Arabic character forms. In conforms with the SASO recommendations for the Arabic character set.

Alteration of Arabic characters to their correct start, medial, or terminal forms is determined and handled by the terminal. Latin and Arabic can be mixed in the same line, the screen displaying both and without the field being altered. Mode change is handled by control keys.

All aspects of the Arabic/Latin features are controlled by the terminal through a Z80 micro-processor; no special software is required on the host computer.

Ascribe was developed by the al-Mudandis in conjunction with an affiliate, International Computer Systems, London, and is manufactured in the UK by TDS.

Al-Arabi is based upon the TDV 2230, manufactured by the Norwegian Company Tandberg. Arabic features have been added to this terminal. The terminal, which is compatible with the DEC VT100 terminal is now being used throughout the Saudi Arabian National Bank's online system.

Scripto-Print, a full bi-lingual letter quality printer provides output for the Ascribe, while a new version is being introduced for al-Arabi.

Arabic Latin Information Systems (ALIS)

ALIS develops systems and terminals based on the Arabic terminal design of Dr. Hyder. The company has developed bilingual dot matrix printers and computer terminals. Its products include:

- Video terminals based on the ADDS Regent 40; the 1311 simultaneously displaying Arabic and Latin; the 1211, handling both languages but not simultaneously.

TABLE 3. CODAR-U.

b^7	0	0	0	0	1	1	1	1
b^6	0	0	1	1	0	0	1	1
b^5	0	1	0	1	0	1	0	1
b^4 b^3 b^2 b^1	0	1	2	3	4	5	6	7
0 0 0 0	0 NUL	TW_7	SP	0	ـ	ـ	ـ	ظ
0 0 0 1	1 TC_1 (soh)	DC_1	!	1	ـ	ـ	ـ	ع
0 0 1 0	2 TC_2 (stx)	DC_2	"	2	ـ	ـ	ـ	غ
0 0 1 1	3 TC_3 (etx)	DC_3	#	3	ـ	ـ	ـ	ف
0 1 0 0	4 TC_4 (eqt)	DC_4	$	4	ـ	ـ	ـ	ق
0 1 0 1	5 TC_5 (enq)	TC_8 (nak)	%	5	ـ	ـ	ـ	ك
0 1 1 0	6 TC_6 (ack)	TC_9 (syn)	&	6	ـ	ـ	ـ	ل
0 1 1 1	7 BEL	TC_{10} (etb)	'	7	ـ	ـ	ـ	م
1 0 0 0	8 FE_0 (bs)	CAN	(	8	ـ	ـ	ـ	ن
1 0 0 1	9 FE_1 (ht)	EM	)	9	ـ	ـ	ـ	ه
1 0 1 0	10 FE_2 (lf)	SUB	*	:	ـ	ـ	ـ	و
1 0 1 1	11 FE_3 (vf)	ESC	+	;	ـ	ـ	ـ	ي
1 1 0 0	12 FE_4 (ff)	IS_4 (fs)	,	<	ـ	ـ	ـ	ـ
1 1 0 1	13 FE_5 (cr)	IS_3 (gs)	-	=	ـ	ـ	ـ	ى
1 1 1 0	14 SO	IS_2 (rs)	.	>	ـ	ـ	ـ	ة
1 1 1 1	15 SI	IS_1 (us)	/	؟	ـ	ـ	ـ	DEL

code can be allocated for each letter of the alphabet, irrespective of its shape [4].

Requirements for Arabic Terminals Suitable for Library Use

Although a number of different manufacturers produce Arabic/Latin terminals and printers, none, so far, has produced what could be termed a general purpose Arabic terminal with a standardized character set.

With the introduction of CODAR-U/FD, it will make more sense for manufacturers of Arab/Latin terminals to use one of the following two methods which are possible for representing the combined Arabic/Latin character sets:

(a) 7-bit CODAR-U/FD (Arabic) and 7-bit ASCII (Latin), with an appropriate escape sequence to switch between character sets and modes of operation (right-to-left and left-to-right)

(b) 8-bit combined Arabic-Latin (extension of ISO specifications 646 [12], 2022 [13], and 2375 [14]).

The hardware eventually chosen by information specialists for processing information in the Arabic language should also allow for:

(a) Full upper and lower case Latin plus CODAR-U/FD
(b) Microprogrammed elaboration of the form of the Arabic characters according to context, with the automatic display of the appropriate shapes of character endings. This will permit the storage of a unique code per character within the computer while allowing for correct display to the user.
(c) Mixed Arabic/Latin characters anywhere in the Arabic mode.
(d) Character matrix of sufficient resolution to permit easy recognition of individual characters.

TABLE 2. Multiplicity of shapes of Arabic characters.

Initial	Medial	Final	Alone	Value[5]
ا	ـا	ـا	ا	a
بـ	ـبـ	ـب	ب	b
تـ	ـتـ	ـت	ت	t
ثـ	ـثـ	ـث	ث	th
جـ	ـجـ	ـج	ج	j
حـ	ـحـ	ـح	ح	ḥ
خـ	ـخـ	ـخ	خ	kh
د	ـد	ـد	د	d
ذ	ـذ	ـذ	ذ	dh
ر	ـر	ـر	ر	r
ز	ـز	ـز	ز	z
سـ	ـسـ	ـس	س	s
شـ	ـشـ	ـش	ش	sh
صـ	ـصـ	ـص	ص	ṣ
ضـ	ـضـ	ـض	ض	ḍ
طـ	ـطـ	ـط	ط	ṭ
ظـ	ـظـ	ـظ	ظ	ẓ
عـ	ـعـ	ـع	ع	'
غـ	ـغـ	ـغ	غ	gh
فـ	ـفـ	ـف	ف	f
قـ	ـقـ	ـق	ق	q
كـ	ـكـ	ـك	ك	k
لـ	ـلـ	ـل	ل	l
مـ	ـمـ	ـم	م	m
نـ	ـنـ	ـن	ن	n
هـ	ـهـ	ـه، ـة	ه، ة	h
و	ـو	ـو	و	w
يـ	ـيـ	ـي	ي	y

good quality newspaper will have 150, a typewriter 65 to 75, while a telex machine will have 52 characters [4].

Arabic/Latin I/O Devices

Over the last two years, an increasing number of sophisticated users have begun to rely upon computers on a large scale in text processing and in small businesses thus emphasizing the need for better bilingual terminals.

Recently, it has become apparent that standardization of Arabic characters and their coding inside the computer are urgently needed to ensure compatibility of products produced by various manufacturers.

Most existing I/O units suffer from one or more of the following shortcomings [6]:

(1) Internal representation of Arabic code is not in any logical sequence.
(2) No vocalization symbols are represented.
(3) Character shapes are poorly represented and not pleasing to the reader's eye.
(4) Keyboard and keyboard layout are not acceptable and are poorly designed.
(5) Capabilities of the units are very limited.

Towards a Unified Arab Code

Between the years 1976–1982, computer experts from the Arab world convened a series of meetings in an attempt to develop a unified code for the use of Arabic characters in informatics. The most significant of the meetings were:

- June, 1976: IBI (Intergovernmental Bureau of Informatics)—the first step towards open discussion concerning the standardization of the use of Arabic in data processing; also led to the formation of COARIN (Committee on Arabic in Informatics) [7].
- April, 1977: ALECSO (Arab League Educational, Cultural, and Scientific Organization)—an agreement was reached on the adoption of a unified Arabic code and on the number of Arab characters to be represented, as well as the diacritics needed to produce an acceptable Arabic text [8].
- June 1977: COARIN—approved the joint Iraqi-Moroccan proposal for a unified Arabic code for informatics known as CODAR-U, which consisted of a 31 Arabic character set and several optional sets that take care of the vocalization symbols [9]. (See Table 3.)
- June 1980: SASO (Saudi Arabian Standards Organization)—accepted CODAR-U as a starting point toward a unified Arabic code
- September 1980: ALECSO—a modified and improved version of CODAR-U [10].
- April 1982: ALECSO—an agreement was reached to accept the modified version of CODAR-U known as CODAR-U/FD (Unified Arab Code—Final Version) [11]. (See Table 4.)

The importance of using CODAR-U/FD is due to the fact that several Arab countries have worked together for seven years to develop the appropriate set of Arabic characters that can be used in the computer field throughout the Arab world.

The Hydriyya Method

This method, introduced by Professor Hyder at the Department of Computer Science, University of Montreal, Canada, requires that when a key is pressed, the letter is not printed immediately but is stored in the terminal's buffer. When the next character or space is typed, the terminal is then in a position to calculate both the correct shape of the first letter and the other code(s) needed to perform it. The second letter of the word is displayed when the third letter is keyed, and so the process is continued until the word is completed.

Under this method the terminal keyboard includes only the isolated Arabic character set. The proper shape of the character is selected automatically via the Arabic Script Processor (ASP) contained in the operating firmware. ASP will select the shape for every individual character according to its preceding and following character [4].

To print/display the last letter in a word, the terminal checks the type of the preceding letter to decide whether the end letter is to be joined or if it is to be separate. The major advantage of this approach is that one key and

(1) Arabic is written from right to left and most letters in the Arabic word are joined together. Arabic numbers are also written from right to left. However, the most significant digit is found on the left so they do not appear reversed.

(2) The Arabic alphabet consists of 31 characters (28 actually, plus the Hamza which appears as a separate character in the written language, but is rarely used alone, Ta Marbutah which is one of three possible shapes for the letter ta' when it appears as a stand-alone terminal character, and Alif Maqsurah which is represented by the long vowel romanized as in à Mustafa or Rida). Twenty-two of these letters can be joined on both sides and in the process take different shapes depending on their context in a word. The position can be in the beginning of the word, like ـحـ , or in the middle of the word, like ـحـ, or at the end of the word, like ـح . The letter can also be written separately not connected to another letter in the same word, like ب , as in بابا . The final letter, ـبـ, will cause a space within a word. Table 1 shows the Arabic letters, their names and pronunciation. As these figures illustrate, the Arabic character can be divided into junction-line, body and tail. The junction-line never changes shape, the body can change shape depending upon position, and the tail has different shapes which are related to the character to which it is added. In the examples presented in Table 2 it is shown that it is necessary to know the shapes of the preceding character and the two subsequent characters before the shape of a given character can be determined.

(3) The Arabic script is highly calligraphic. When Arabic printing was introduced early in the eighteenth century, the shapes of Arabic letters were an imitation of handwriting which was calligraphic. Thus, at first, printing was an imitation of calligraphy, even if a little simplified. As literacy spread, the opposite became true; writing became a simplification of printing [3]. This process continued, and still continues. Because of the still-powerful traditions of calligraphy, the users of the Arabic language tend to resist attempts to compromise aesthetics and quality of print for the sake of mechanical efficiency.

(4) Besides the alphabet character set there are at least five vowels. These are optional in writing, used only when it is not possible to guess the exact word from the context. They are used in conjunction with other letters printed above or below the letter.

While the first problem is not serious, the second and third are. Vocalization symbols are essential for the semantics of Arabic sentences and words. These symbols are not customary in commercial type applications. In more complex applications in the field of information storage, search and retrieval, artificial intelligence, and language theory it is impossible to ignore them.

Thus, while the letters of the Latin alphabet have only one form, the sole exception being the use of capitals, this is not the case with Arabic script. Some characters appear in four, or possibly more, different shapes. Two characters (Alif and Lam) have special shapes whenever the Alif follows the Lam. There can be up to 900 letter shapes printed in calligraphic style. Currently, a well written Arabic book will have about 450 shapes, a fairly

TABLE 1. Arabic letters, their names, and pronunciations.

1	ء	hamza:	glottal stop	17	ض	dad:	ḍ emphatic d
2	ا	alif:	a	18	ط	ta':	ṭ emphatic T
3	ب	ba':	b	19	ظ	dha':	ẓ emphatic dh
4	ت	ta':	t	20	ع	'ayn:	(no equivalent)
5	ة	ta'maqsurah:	t or h	21	غ	ghayn:	gh French Pronounciation of r
6	ث	tha':	th as in thin	22	ف	fa':	f
7	ج	jim:	j as in joke	23	ق	qaf:	q uvular k
8	ح	ha':	ḥ Spanish pronounciation of ha	24	ك	kaf:	k
9	خ	kha':	kh German pronounciation of ch	25	ل	lam:	l
10	د	dal:	d	26	م	mim:	m
11	ذ	thal:	dh as in this	27	ن	nun:	n
12	ر	ra':	r	28	ه	ha':	h
13	ز	za':	z	29	و	waw:	w
14	س	sin:	s	30	ي	ya':	y as in yes
15	ش	shin:	sh	31	ى	alif-maqsura:	a
16	ص	sad:	ṣ emphatic s				

(b) Vowel Points

32	ـَ	fathah:	(a)
33	ـُ	dammah:	(u)
34	ـِ	kasrah:	(i)
35	ـّ	shaddah:	(doubling)
36	ـْ	sukun:	(stop)