ECONOMIC AND SOCIAL COMMISSION FOR WESTERN ASIA

HARMONIZATION OF ICT STANDARDS RELATED TO ARABIC LANGUAGE USE IN INFORMATION SOCIETY APPLICATIONS

United Nations

Distr.
GENERAL
30 April 2003
E/ESCWA/ICTD/2003/5
ORIGINAL: ENGLISH

ECONOMIC AND SOCIAL COMMISSION FOR WESTERN ASIA

HARMONIZATION OF ICT STANDARDS RELATED TO ARABIC LANGUAGE USE IN INFORMATION SOCIETY APPLICATIONS

United Nations
New York, 2003

Bibliographical references have, wherever possible, been verified.

The views expressed in this paper do not necessarily reflect the views of the United Nations Secretariat.

03-0357

Summary

The Internet is rapidly playing a more pivotal role in society, and access to electronically transmitted information and to the growing online marketplace is proving to be essential for the economic development of communities and nations in the Middle East region. Despite the obstacles related to the use of Arabic on the Internet, advances are being made. The number of Arabic Internet users grew from an estimated 4.4 million in March 2002 to 5.5 million in March 2003. Moreover, while those who utilize the Arabic language constitute over 5 per cent of the non-English population worldwide, the corresponding percentage of Internet users is only slightly over 1 per cent. This discrepancy can be very much attributed to the absence of unified information and communications technology (ICT) standards related to the use of the Arabic language in information society applications. Arabic usage on the Internet has also been negatively affected by the diversity of standards for specific areas, namely, the multiplicity of character sets.

In this context, the study highlights the urgent need to formulate and adopt standards for the use of Arabic language in ICT. It also stresses the importance of Arabizing the Internet, and discusses tools that can facilitate this, including open source software and Arabic domain names.

Accordingly, the study examines existing ICT standards related to the Arabic language, and reviews Arab and international organizations. Classification of Arabic language standardization issues, including character encoding, e-mail, web browsing, speech recognition/synthesis, and searching and indexing on the Internet are examined and the increasing utilization of Unicode, which is also known as the Universal Multiple-Octet Coded Character Set (UCS), to standardize the exchange of Arabic on the Internet is also detailed. Furthermore, the study emphasizes the need to ensure that existing ICT standards are translated into Arabic.

The study also considers legislative and regulatory issues for information and knowledge management (IKM) in relation to ICT terminology and Internet security, and highlights the need to do the following: (a) unify ICT terminologies in all ICT-related Arabic language standards; (b) form an Internet Best Practices Committee for the Arabic language; (c) address security issues on the Internet; and (d) standardize digital preservation strategies. The study concludes by proposing an action plan to improve the utilization of Arabic on the Internet.

d

CONTENTS

Page

Summary iii
Abbreviations vii

Chapter

I. ARABIZATION IN THE DIGITAL WORLD 1

A. Arabic calligraphy 3
B. Arabization of the Internet 8
C. Open source software 10
D. Issues pertaining to Arabic domain names 15
E. Importance of Arabic ICT standards 19
F. Existing standards 22

II. CLASSIFICATION OF ARABIC LANGUAGE STANDARDIZATION ISSUES 25

A. Translation of web pages 25
B. Internet 26
C. Languages 29
D. Hardware 30
E. Applications 31

III. PRIORITY ISSUES IN ARABIC LANGUAGE STANDARDIZATION 33

A. Arabic character sets and their standards 33
B. Arabic display issues 33
C. Arabic e-mail 34
D. Arabic Web browsing 35
E. Arabic searching and indexing 37
F. Speech technology 38
G. Optical character recognition 41

IV. LEGISLATIVE AND REGULATORY ISSUES FOR INFORMATION AND KNOWLEDGE MANAGEMENT 43

A. ICT terminology 43
B. Standards for the utilization of Arabic language on the Internet 43
C. Security issues 44
D. Digital preservation strategies 48

V. CONCLUSION AND RECOMMENDATIONS 49

LIST OF TABLES

1. Number of people online in each language zone 2
2. Arabic language online population and Internet users in relation to Asian, non-English and world online populations and Internet users 2
3. Internet penetration in the Arab countries 16
4. Results of a survey pertaining to Arabizing international domain names 17

CONTENTS (*continued*)

Page

LIST OF FIGURES

I. Distribution of the online language population 2
II. The different shapes of the 28-letter Arabic alphabet 4
III. Differentiation of Arabic letters through the use of diacritics 5
IV. Arabic, Phoenician and other alphabets 5
V. Al Khalil ibn Ahmad al Farahidi *Tashkeel* system, including six diacritics 6
VI. Measuring system of Ibn Muqlah based on a circle with a diameter that equals the height of the letter *Alef* 7
VII. The Arab world 8
VIII. Domain name system operations for an internationalised domain name system 19

LIST OF BOXES

1. The art of Arabic calligraphy: language and script 3
2. The art of Arabic calligraphy: a brief history 6
3. Arabic calligraphy styles 7
4. Arabization 9
5. Evolution of Linux and the benefits of its Arabization project 11
6. Arabic domain names 15
7. iDNS solutions for internationalizing DNSs 19
8. Internationalization problems for the Arabic language 21
9. Standards pertaining to Arabic 23
10. Standards pertaining to information technology security 24
11. Computers and characters 26
12. Internationalization of the Web 29
13. Character set definition 33
14. Representing Arabic text and displaying bi-directional text 34
15. Encoding e-mail message body and representing body and header character sets 35
16. Relevant internationalization features for Arabic 36
17. Other Arabization techniques 37
18. ICT vulnerability 45
20. ICT security elements 46

ANNEXES

I. Unicode 52
II. Arabic code tables 55
III. Arabic and international organizations 60

ABBREVIATIONS AND ACRONYMS

AAG	Arab Advisors Group
ADN	Arabic domain name
AIDMO	Arab Industrial Development and Mining Organization
AIDMO-CSM	AIDMO-Center of Standardization and Metrology
AIDNA	Arab Internet and Domain Name Association
AGIP	Abu-Ghazaleh Intellectual Property
AINC	Arabic Internet Names Consortium
AKMS	Arab Knowledge Management Society
ALECSO	Arab League Educational, Cultural and Scientific Organization
ALC	Arabic Linguistic Committee
ALUG	Arab Linux User Group
Arabcin	Arab Club for Information
ARIFONET	Arab Industrial Information Centers
ARSO	African Regional Organization for Standardization
ASIP	Arab Society for Intellectual Property
ASM	Arabic software map
ASMO	Arab Standards and Metrology Organization
ASP	application service provider
ASR	automatic speech recognition
ASTM	American Society for Testing and Materials
AUCBM	Arab Union for Concrete and Building Materials
BC	Business Constituency
BSMD	Bahrain Directorate of Standards and Metrology
C2M	Consulting to Management
CCS	coded character set
ccTLD	country code top-level domain
CDSA	common data security architecture
DES	data encryption standard
DGSM	Directorate General for Specifications and Measurements of Oman
DNS	domain name system
DNSO	Domain Name Supporting Organization
DOS	Disk Operating System
DSS/DSA	digital signature standard/digital signature algorithm
EAD	encoded archival description
ECMA-ATF	European Computer Manufacturing Association, Arabic Task Force
EDI	electronic data interchange
EFI	electronic form interchange
EOS	Egyptian Organization for Standardization and Quality Control
ESMTP	extended simple mail transfer protocol
ESCWA	Economic and Social Commission for Western Asia
ETSI	European Telecommunications Standards Institute
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
FIPS	Federal Information Processing Standards
GCC	Gulf Cooperation Council
GSMO	Gulf Standards and Metrology Organization
gTLDs	generic TLDs
HMD	head mounted display
HTML	hypertext markup language
HTTP	hypertext transfer protocol
IANA	Internet assigned number authority
ICANN	Internet Corporation for Assigned Names and Numbers
ICC	International Chamber of Commerce
ICOSM	Iraqi Central Organization for Standardization and Metrology

ABBREVIATIONS AND ACRONYMS (*continued*)

ICT	information and communications technology
iDNS	internationalized DNS
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IEEE-SA	IEEE Standards Association
IETF	Internet Engineering Task Force
IFAS	International Federation for the Application of Standards
IKE	Internet key exchange
IKM	information and knowledge management
IP	Internet Protocol
IPC	Intellectual Property Constituency
ISAKMP	Internet Security Association and Key Management Protocol
ISO	International Organization for Standardization
ISOC	Internet Society
ISP	Internet service provider
IT	information technology
ITU	International Telecommunication Union
JISM	Jordanian Institution for Standards and Metrology
KEA	Key Exchange Algorithm
KSMD	Kuwait Standards and Metrology Department
LDC	Linguistic Data Consortium
LIBNOR	Lebanese Standards Institution
LUG	Linux user groups
MARC	machine-readable cataloguing
MB	megabytes
MEES	Middle East Economic Survey
MHTML	Multilingual-HTML
MIME	Multipurpose Internet mail extensions
MINC	Multilingual Internet Names Consortium
NLP	natural language processing
NSB	National Standardization Body
OAIS	Open Archival Information System
OCR	optical character recognition
ODS	Official Document System of the United Nations
OIML	International Organization of Legal Metrology
OS	operating system
OSS	open source software
PC	personal computer
PSSE	Palestinian Standards and Specifications Establishment
QC/QA	quality control/quality assurance
QDSMCP	Qatar Deptartment of Standards, Measurements and Consumer Protection
R&D	research and development
RACE	row-based ASCII compatible encoding
RFC	Requests for Comments
RITSEC	Regional Information Technology and Software Engineering Center
RSA	Rivest-Shamir-Adelman
SASMO	Syrian Arab Organization for Standardization and Metrology
SASO	Saudi Arabian Standards Organization
SaudiNIC	Saudi Network Information Center
SCC10	Standards Coordinating Committee 10
SGML	standard generalized markup language
SHS/SHA	secure hash system/secure hash algorithm
SMTP	simple mail transfer protocol

ABBREVIATIONS AND ACRONYMS (*continued*)

SNNs	sub-neural-networks
SSUAE	Directorate of Standardization and Metrology of the United Arab Emirates
TAGI	Talal Abu-Ghazaleh International Society
TC-8	Technical Committee 8 (AIDMO-CSM)
TLD	top-level domain
TTS	text-to-speech
UCS	Universal Multiple-Octet Coded Character Set (Unicode)
UNBISnet	United Nations Bibliographic Information System
UNCAPS	United Nations Shared Cataloguing System
UNCITRAL	United Nations Commission on International Trade Law
UNCTAD	United Nations Conference on Trade and Development
URI	uniform resource identifier
URL	uniform resource locator
UTF	Unicode transfer format
W3C	World Wide Web Consortium
WIPO	World Intellectual Property Organization
WTO	World Trade Organization
WWW	World Wide Web
XML	EXtensible markup language

Notes: References to dollars ($) are to United States dollars, unless otherwise stated.

Certain abbreviations are not explained in the text owing to their widespread usage in this form.

I. ARABIZATION IN THE DIGITAL WORLD

Despite many obstacles, Arabic is increasingly being used on the Internet. However, further expansion in this area is being hindered by the lack of unified standards, particularly in the field of character sets. Other problems regarding the use of Arabic on the Internet were uncovered by a survey carried out in Saudi Arabia at the beginning of 1997, which identified a number of obstacles to greater Internet usage in the Arab world, namely, weak telecom infrastructure; lack of Arabic content on the Internet; and lack of Arabic Internet access programmes for the Web and electronic (e)-mail.[1]

The use of Arabic on the Internet can be improved by providing support in the fields of content, transport, and client and server processing.[2] Arabic textual content relates to the representation of data and its formatting through Internet standards, namely, hypertext markup language (HTML) for World Wide Web (WWW) or the Web pages, and Requests for Comments (RFC) 822 and multipurpose Internet mail extensions (MIME) for e-mail messages. Hypertext transfer protocol (HTTP) is the transport protocol for the Web and simple mail transfer protocol (SMTP) is the transfer protocol for e-mail. Client processing refers to the generation, display and interaction with Arabic text, while server processing involves storing, processing, searching and providing Arabic content.

Perhaps the most intrinsic difficulty for users of Arabic on the Internet is the multiplicity of character sets. Arabic text is cursive, and the written form of Arabic letters varies according to their position in the word. Moreover, Arabic letters are written right-to-left whereas Arabic numbers are written left-to-right. The difference in directionality between Arabic and Latin text, and the frequent need to combine Arabic and Latin text on the same line necessitates a system that can handle bi-directional text. Other problems exist with regard to the searching and indexing of Arabic texts, representation and transport, and the display of Arabic features.

Solutions are emerging, and browsers and mail programmes have been building on Internet standards such as MIME. Unicode, which is also known as UCS, is also facilitating the exchange of Arabic on the Internet (see annex I). Other interim solutions are frequently used, namely, encoding text as graphics and relying on ad hoc rules in Web servers to ascertain the Arabic capabilities of browsers and to send information accordingly. The increasing internationalization of Internet protocols is also an encouraging development.[3]

Global Internet statistics categorized by language are available on the Internet.[4] The estimated statistics for March 2002 and March 2003 for the number of people online in each language zone are highlighted in table 1, classified by language rather than by country. This classification has been utilized on the basis that people speaking the same language tend to form their own online community, no matter what country they happen to live in.

In March 2003, the estimated number of Arabic language Internet users amounted to 2.7 per cent of the total Asian language Internet users, 1.3 per cent of total non-English Internet users, and 0.8 per cent of total world Internet users. However, users of the Arabic language constituted 18.1 per cent of the total Asian language population at that time, 5.2 per cent of the total non-English language population, and 4.7 per cent of the total world population (see table 2). Therefore, while the Arab population constituted 18.1 per cent of the Asian population, they represented only 2.7 per cent of Asian language Internet users. Similarly, despite the fact that the Arab population accounted for approximately 5 per cent of the population of the world, it represented less than 1 per cent of Internet users worldwide.

---

[1] B. Al-Badr, "Using the Internet in Arabic: problems and solutions", Internet Summit 1998 (INET '98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

[2] B. Al-Badr, "Standards for supporting Arabic on the Internet", *Proceedings of the First King Saud University Workshop on Internet Arabization*, Riyadh, 18 May 1997 (in Arabic).

[3] H. Alvestrand, "IETF policy on character sets and languages", (The Internet Society, January 1998). Available at: http://www.ietf.org/rfc/rfc2277.txt and http://www.faqs.org/rfcs/rfc2277.html.

[4] Global Reach, "Global Internet statistics (by language)". Available at: http://www.glreach.com/globstats/index.php3.

TABLE 1. NUMBER OF PEOPLE ONLINE IN EACH LANGUAGE ZONE

	March	Internet access (millions)	World online population (percentage)	2004 (estimated in millions)	Total population (millions)	Gross domestic product (millions of dollars)	World economy (percentage)	Net hosts
English	2002	228	40.2		567	13 812	33.4	
	2003	238	35.2	280	508			
Non-English	2002	339	59.8		5 633	27 590	66.6	
	2003	440	64.8	680	5 822			
Arabic	2002	4.4	0.8		300	678	1.6	95
	2003	5.5	0.8	7	300	678	1.6	1 592
European languages (non-English)	2002	192.3	33.9		1 218	12 550	30.3	24 529
	2003	238	35.1	328	1 218	12 968	33.9	24 529
Asian languages	2002	146.2	26.1		1 658			10 440
	2003	201.7	29.7	263	1 658			10 440
World	2002	560			6 200	41 400		
	2003	648.7		940	6 330	41 400		

Source: Adapted from Global Reach, "Global Internet statistics (by language)". Available at: http://www.glreach.com/ globstats/index.php3.

Note: Two dots (..) indicate that data are not available or is not separately reported.

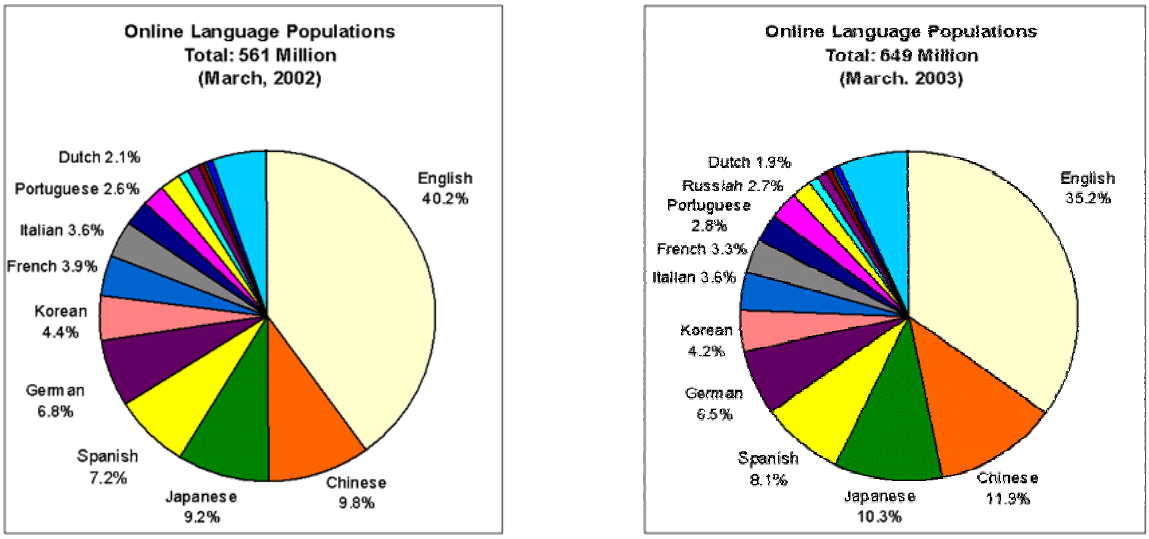
TABLE 2. ARABIC LANGUAGE ONLINE POPULATION AND INTERNET USERS IN RELATION TO ASIAN, NON-ENGLISH AND WORLD ONLINE POPULATIONS AND INTERNET USERS (*Percentage*)

	Asian		Non-English		World	
Language	2002	2003	2002	2003	2002	2003
Arabic Internet users	3	2.7	1.3	1.3	0.8	0.8
Arabic population	18.1	18.1	5.3	5.2	4.8	4.7

Source: ESCWA, compiled from various sources.

The distribution of the online language population, which totalled an estimated 561 million in March 2002 and an estimated 649 million in March 2003, is illustrated in figure I.

Figure I. Distribution of the online language population



Source: Adapted from Global Reach, "Global Internet statistics (by language)". Available at: http://www.glreach.com/ globstats/index.php3.

Note: The 2002 chart is no longer available online.

The English language is widely used on the Web, with recent estimates indicating that approximately 70 to 80 per cent of existing web pages are written in English. One of the reasons for the prevailing position of the English language includes the high percentage of Internet users in the United States of America. Indeed, whilst it is true that the English language is predominant on the Internet, there is evidence that the number of non-English speaking Internet users is steadily increasing as penetration rates in non-English speaking countries continue to grow.

The number of Internet users who speak a particular language determines the amount of content that is available in that language. In other words, the existence of more content in a particular language encourages speakers of that language to go online. A certain percentage of these new users will go on to produce their own content, which will in turn increase the overall content available in that language, thereby forming a positive feedback loop.

According to the Computer Industry Almanac,[5] the number of Internet users surpassed 530 million in 2001 and will continue to increase steadily in the next few years. The majority of growth can be witnessed in Asia, Latin America and parts of Europe. By the end of 2005 the number of worldwide Internet users is expected to exceed 1 billion and an increasing portion of Internet users will be using wireless devices such as Web-enabled cell phones and personal digital assistants (PDAs). The wireless devices will complement personal computer (PC) Internet access for most users in developed countries. In countries with low Internet penetration, wireless devices will be the primary Internet access devices.

A. ARABIC CALLIGRAPHY

Arabic belongs to the group of Semitic alphabetical scripts that use letters to represent consonants, and on an optional basis, diacritics to indicate vowels. Various aspects of Arabic calligraphy are examined in boxes 1, 2 and 3.

Box 1. The art of Arabic calligraphy: language and script

"The earliest-known alphabet to mankind was the North Semitic, which developed around 1700 B.C. in Palestine and Syria. It consisted of 22 consonant letters. The Arabic, Hebrew, and Phoenician alphabets were based on this model. Later, around 1000 B.C., the Phoenician alphabet was itself used as a model by the Greeks, who added letters for vowels. Greek in turn became the model for Etruscan (c. 800 B.C.), whence came the letters of the ancient Roman alphabet, and ultimately all Western alphabets.

"The North Arabic script, which eventually prevailed and became the Arabic script of the *Quran*, relates most substantially and directly to the Nabatian script, which was derived from the Aramaic script. Old Aramaic, the language of Jesus and the Apostles, dates from the 2nd millennium B.C., and some dialects of which are still spoken by tiny groups in the Middle East. Arabic script still shares with Aramaic the names of the alphabet letters (*Alef, Jeem, Dal, Zai, Sheen*, etc.); similar graphic representation for phonetically similar letters (*Sad, Dad, Ta and Tha*, etc.); connections of letters in the same word and several forms of each letter depending on its location in the word, except for letters that cannot be connected to the letters which come after them (*Alef, Dal, Raa, Waw*).

"The Arabic Alphabet has 28 letters (see figure II). The shape of these letters changes depending on their position in the word, whether isolated; in the beginning of the word (initial); in the middle (medial); or at the end (final). Several letters in the Arabic alphabet share the same shape (see figure III), and are differentiated only by the number and placement of (one, two, or three) dots on the letters. Of the basic 18 letter shapes, 2 are used for three letters, 6 are used for two letters, and the remaining 10 are used for one letter each. The Arabic and Phoenician alphabets, along with several other alphabets (see figure IV) such as Hebrew and Aramaic, are based on an early model called the North Semitic. The Phoenician alphabet was adapted by the Greeks, then the Etruscans and Romans, and eventually became the Western alphabet as we know it today.

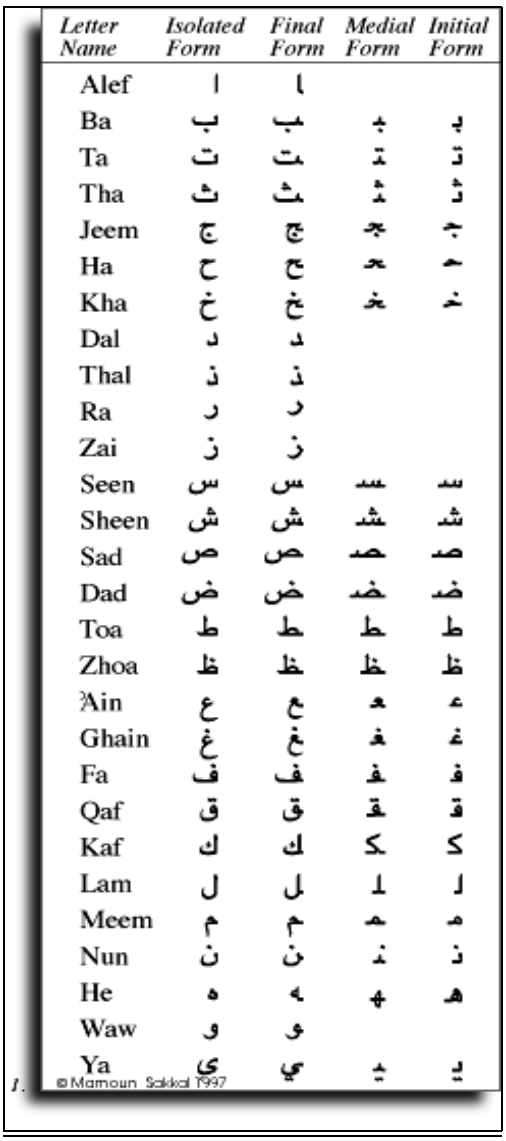
[5] Available at: http://www.c-I-a.com/.

Box 1 (*continued*)

"With the spread of Islam, the Arabic alphabet was adapted by several non-Arab nations for writing their own languages. In Iran, Arabic letters were used to write Farsi, with the addition of four letters to represent the phonetics that did not exist in Arabic: p, ch, zh and g. The Ottoman Turks used the Arabic alphabet until 1929 and added still another letter. This alphabet was also used to write other Turkish languages and dialects, such as Kazakh, Uzbek, etc. Several other languages used the Arabic alphabet at one time or another, including Urdu, Malay, Swahili, Hausa, Algerian Tribal, and others. From its simple and primitive early examples of the 5th and 6th century A.D., the Arabic alphabet developed rapidly after the rise of Islam in the 7th century into a beautiful form of art. The main two families of calligraphic styles were the dry styles, called generally the *Kufic*, and the soft cursive styles, which include *Naskhi*, *Thuluth*, *Nastaliq* and many others".

Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabicCalligraphy.html.

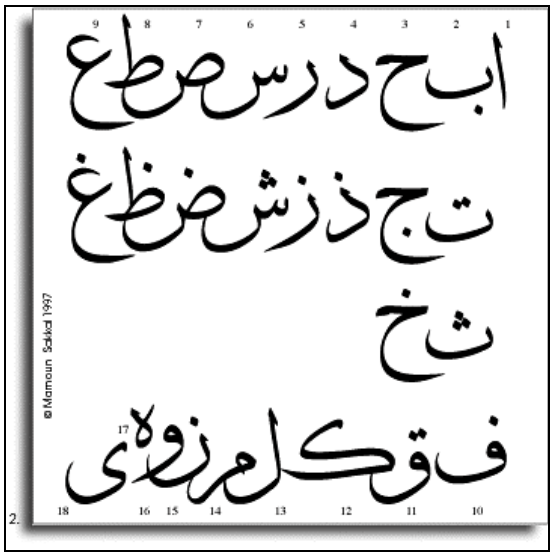
Figure II. The different shapes of the 28-letter Arabic alphabet



Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabicCalligraphy. html.

Note: Certain Arabic alphabet letters share the same shape, and are only differentiated by the number and placement of diacritics or dots on the letters.

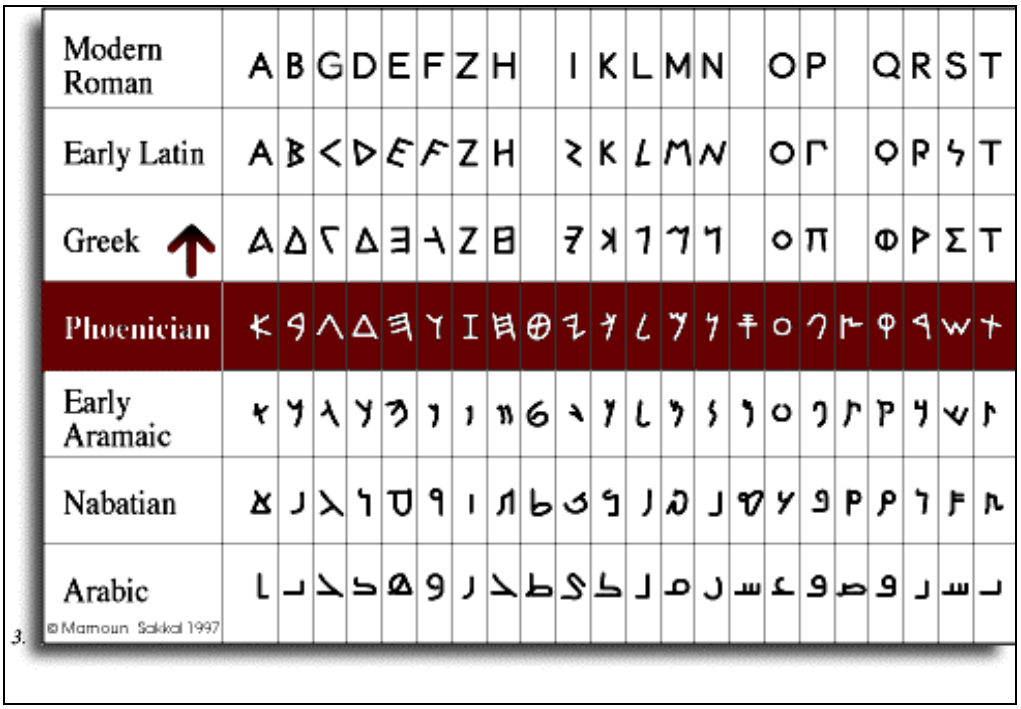
Figure III. Differentiation of Arabic letters through the use of diacritics



Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabicCalligraphy. html.

Note: Certain Arabic alphabet letters share the same shape, and are only differentiated by the number and placement of diacritics or dots on the letters.

Figure IV. Arabic, Phoenician and other alphabets



Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabicCalligraphy.html.

Box 2. The art of Arabic calligraphy: a brief history

"The North Arabic script, which was influenced by the Nabatian script, was established in north-eastern Arabia and flourished in the 5th century among the Arabian tribes who inhabited Hirah and Anbar. It spread to Hijaz in western Arabia, and its use was popularized among the aristocracy of Quraysh, the tribe of the Prophet Muhammad, by Harb ibn Ummayyah.

"Although early Arabic sources mention several calligraphic styles in reference to the cities in which they were used, they generally fit into two broad categories with some minor variations, these are the "dry styles," the early predecessors of *Kufic*, and the "moist styles", the early predecessors of the cursive family or scripts.

"With the increasing number of non-Arab Muslims, there was a greater need for facilitating reading and learning of Arabic. Since several letters of the Arabic alphabet share the same shapes, and since vowels are not clearly indicated, some reform was needed to avoid confusion, and a system of *Naqt* or *I'jam* (letter-pointing), and *Tashkeel* (vowel indication) was developed.

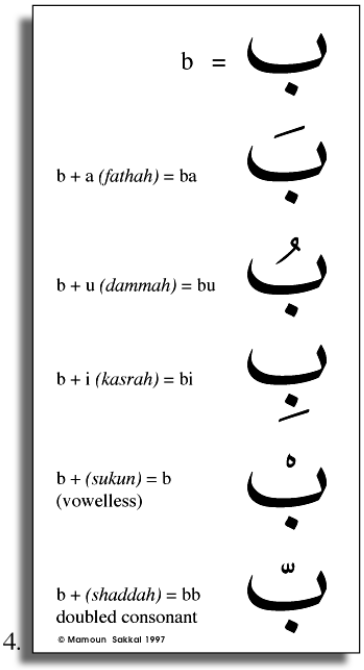
"Abul Aswad al Du'ali (d. 688) was the legendary founder of Arabic grammar, and is credited with inventing the system of placing large colored dots to indicate the *Tashkeel*. It was used with the *Kufic* scripts, but proved to be somewhat cumbersome to use with smaller scripts, or in ordinary writing.

"The Ummayad governor al Hajjaj ibn Yusuf al Thaqafi enforced a uniform system to distinguish letters by using dots, which he asked two of al Du'ali's students to codify. Al Khalil ibn Ahmad al Farahidi (d. 786) devised a *tashkeel* system to replace Abu al Aswad's (see figure V). His system was universally used since the early 11th century, and six diacritical marks to indicate the small vowels attached to Arabic letters: *Fathah* (a), *Dammah* (u), *Kasrah* (i), *Sukun* (without vowels), *Shaddah* (double consonant), and *Maddah* (vowel prolongation) which is applied to the *Alef*".

The measuring system of Ibn Muqlah is based on a circle with a diameter that equals the height of the letter Alef (see figure VI). It controls the correct proportions of the letters by comparing them to the circle, and by diagonal dots written with the calligraphy pen.

Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabicCalligraphy.html.

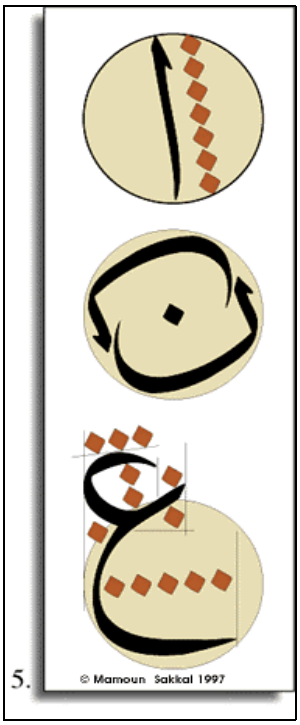
Figure V. Al Khalil ibn Ahmad al Farahidi *tashkeel* system, including six diacritics



4.

Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabic Calligraphy.html.

Figure VI. Measuring system of Ibn Muqlah based on a circle with a diameter that equals the height of the letter *alef*



Source: Mamoun Sakkal, "The art of Arabic calligraphy". Available at: http://www.sakkal.com/ArtArabic Calligraphy.html.

Box 3. Arabic calligraphy styles

The major styles of Arabic calligraphy are reviewed below.

(a) *Deewani*

"*Deewani* script is an Ottoman development parallel to *Shikasteh* (broken style). The script was largely developed … in the late 15th century from the Turkish/Persian Ta'liq. *Deewani* reached its zenith in the 17th century … [and] became a favorite script for writing in the Ottoman chancellery. *Deewani* is excessively cursive and highly structured with its letters undotted and unconventionally joined together. It uses no vowel marks".

(b) *Kufi*

"*Kufi* was the dominant priestly script in early times. It was created after the establishment of the two Muslim cities of Basrah and Kufah in the second decade of the Islamic era (8th century A.D.). The script has specific proportional measurements, along with pronounced angularity and squareness. It became known as *al-Khat al-Kufi* (*Kufi* script)".

(c) *Naskh*

"*Naskh* was one of the earliest scripts to evolve. It gained popularity after being redesigned by the famous calligrapher Ibn Muqlah in the 10th century. Because of Ibn Muqlah's comprehensive system of proportion, Naskh style displays a very rhythmic line.

"*Naskh* later was reformed by Ibn al-Bawaab and others into an elegant script worthy of the *Qur'an*—and more *Qur'ans* have been written in Naskh than in all the other scripts together. Since the script is relatively easy to read and write, *Naskh* appealed particularly to the general population.

"*Naskh* is usually written with short horizontal stems—and with almost equal vertical depth above and below the medial line. The curves are full and deep, the uprights straight and vertical, and the words generally well spaced".

Box 3 (*continued*)

(d) *Riq'a*

"This script, also called *Ruq'ah* (small sheet), evolved from *Naskh* and *Thuluth*. Although *Riq'a* has a close affinity with *Thuluth*, *Riq'a* developed in a different direction and became simplified. The geometric forms of the letters are similar to those of *Thuluth* but are smaller with more curves. *Riq'a* is rounded and densely structured with short horizontal stems. …

"*Riq'a* was one of the favorite scripts of Ottoman calligraphers and underwent many improvements at the hand of Shaykh Hamdullah al-Amasi. Later, *Riq'a* was revised by other calligraphers and went on to become the most popular and widely used script. Today, *Riq'a* is the preferred script for handwriting throughout the Arab world".

(e) *Ta'liq*

"*Ta'liq* (hanging) script, also called Farsi, is an unpretentious cursive script apparently in use since the early 9th century. … The script is currently in great favor with Arabs, and it is the native calligraphic style among the Persian, Indian, and Turkish Muslims.

"The Persian calligrapher Mir Ali Sultan al-Tabrizi developed from *Ta'liq* a lighter and more elegant variety, which came to be known as *Nasta'liq*. However, Persian and Turkish calligraphers continued to use *Ta'liq* as a monumental script for important occasions. The word *Nasta'liq* is a compound word derived from *Naskh* and *Ta'liq*".

(f) *Thuluth*

"*Thuluth* script was first formulated in the 7th century during the Umayyad caliphate, but it did not develop fully until the late 9th century. The name means 'a third' - perhaps because of the proportion of straight lines to curves. … Though rarely used for writing the Holy *Qur'an*, *Thuluth* has enjoyed enormous popularity as an ornamental script for calligraphic inscriptions, titles, headings, and colophons. … *Thuluth* is known for its elaborate graphics and remarkable plasticity".

Source: Islamic Arts and Architecture Organization, "Arabic calligraphy". Available at: http://www.islamicart.com/main/calligraphy/styles/.

It is worth noting that advances in Arabic calligraphy over the past few decades have not taken into account its applicability to ICT systems, which has hindered the utilization of the Arabic language in early printing systems and more currently, in ICT systems.

B. Arabization of the Internet

According to recent estimates, Arabic is the mother tongue of more than 300 million people in 22 Arab countries (see figure VII). While many Arabs prefer to use English or French on the Internet, the majority of Arabs, particularly those in Bahrain, Egypt, Kuwait, Oman, Qatar, Saudi Arabia, the Syrian Arab Republic and the United Arab Emirates use Arabic. Therefore, digital Arabization will increase the number of Arabic language ICT users (see box 4).

Figure VII. The Arab world

Source: Arabic Internet Names Consortium (AINC). Available at: http://www.ainc.tn/arabmap.gif.

Box 4. Arabization

The transformation of the Arab world into an inclusive knowledge based society that is open to all organizations and peoples in the region can be facilitated by ensuring that the Internet can be accessed and used in Arabic. In this context, progress is being made in the following areas.

(a) *Arabization of the Internet*

The Internet Corporation for Assigned Names and Numbers (ICANN), which is partially responsible for the regulations that govern business through the Internet, and which also controls domain names and grants licences to the companies that register domain names, has been exerting efforts to improve the Arabization of the Internet.

The Arab Knowledge Management Society (AKMS) has also been endeavouring to improve the Arabization of the Internet. It established the Arab Internet and Domain Name Association (AIDNA) with the objectives of gaining access to the knowledge world and developing an Arabic Internet; consolidating relationships and cooperation among holders of electronic addresses and domain names; and disseminating Arab culture. The Arab Club for Information (Arabcin) has also been playing an important role in this area by encouraging those interested in developments to share information and contribute to a unified data bank.

(b) *Suitable environment*

The Arab e-business initiative and the Muscat Declaration were both designed to encourage Arab Governments to collaborate in facilitating a suitable environment for a the functioning of a digital information society, and to foster innovation and develop relevant business, technical and legal programmes.

(c) *Publications*

While much needs to be achieved in this area, an increasing number of publications with relevant Arabic content are being produced.

Source: Adapted from: Talal Abu-Ghazaleh International Society (TAGI) and AKMS, "Towards an Arab knowledge society", a joint report, April 2001.

Many organizations, including the Arab Knowledge Management Society (AKMS) and the Talal Abu-Ghazaleh International Society (TAGI), have been endeavouring to Arabize the Internet by enabling end-users to register their Internet addresses using Arabic characters that correspond to ".com", ".net", ".org", ".biz", ".aero", ".coo", ".info", ".museum", ".name", and ".pro" (see annex III). The development of multilingual domain names is based on the need to eliminate the language barrier that prevents the non-native English speaking population from using the Internet (see chapter I, section D). These types of initiatives enable the Arab Internet community to avoid the problems associated with transliteration.

Arabic software that is available on the Internet includes the Arabic software map,[6] which is written in LaTeX and translated into HTML using the LaTeX2HTML translator.[7] The Arabic Software Digest web site is a good source of information concerning Arabic software for communication/networking, publishing, graphics, translation, database, spreadsheet and operating systems (OSs).[8]

---

[6] Isam G. Ishaq, "Arabic script software map". Available at: http://user.cs.TU-Berlin.DE/~ishaq/arabic/asm/asm.html.

[7] Nikos Drakos, "All about LaTeX2HTML". Available at: http://cbl.leeds.ac.uk/nikos/tex2html/doc/latex2html/latex2html.html.

[8] Available at: http://www.gy.com/www/ar.htm.

Arabic browser/e-mail software is produced by the following companies:

(a) *Alis Technologies*[9]

Alis Technologies produces a multilingual web browser that is capable of supporting over 90 languages, including Arabic. Their browser, Tango, supports user interfaces and help-files in many languages. Their products can be evaluated by downloading beta versions of their browser and web creator.

(b) *Sakhr Software*[10]

Sindbad, the Sakhr plug-in for Netscape 3.0, supports Arabic and can be evaluated (after installation) by visiting the Sakhr web site. Sakhr pioneered an Arabization layer for Latin versions of Windows, which provides Arabic support and the correct handling of diacritics. While this Arabization layer lacks a degree of stability, the stand alone products of the company are fairly impressive.

(c) *Accent Technologies*[11]

Accent Technologies, formerly known as Kivun Computers/Accent Software International, develops multilingual programmes for Windows. In 1991, Accent Technologies and Microsoft developed the technology currently used in bi-directional versions of Windows that enables users to enter characters into a document right-to-left and left-to-right. However, while these versions do not provide adequate support for Arabic, a multilingual browser, which supports Arabic is available.

C. OPEN SOURCE SOFTWARE[12]

This section highlights the capabilities of open source.

"What is open source?

(a) "It's the ability to dissect source code line-by-line;
(b) It's the ability to copy/modify building blocks of programs;
(c) It's the ability to distribute modified/personalized code at will;
(d) It's knowing that millions of people scrutinize the code daily;
(e) It's knowing that bugs and vulnerabilities are addressed immediately;
(f) It's security without "blind trust" and/or "promises";
(g) It's knowing that your intellectual property will be maintained."[13]

Arabic can be used in the same manner as Spanish, German and Japanese on open source software (OSS) such as Linux, which is reviewed in greater detail below. EXPLAIN Linux enables programming and documentation to be carried out in English, and provides a user interface that is Arabic-capable. Linux is emerging as a viable OS for businesses all over the world, and has been gaining popularity in the Middle East, where it has not only been adopted by major universities in Bahrain, Egypt, Jordan and Saudi Arabia, but is also being utilized in banks and oil and gas companies in Gulf Cooperation Council (GCC) countries.[14] The Arabization of Linux will further improve this growth. IBM recently completed the Arabization of the

[9] Available at: http://www.alis.com.

[10] Available at: http://www.sakhr.com.

[11] Available at: http://www.accent-technologies.com/.

[12] The history of OSS from 1984 to 1999 is provided in ESCWA, *New Technologies for Enhancing Competitiveness and Productivity in Selected Sectors*, annex IV (E/ESCWA/TECH/2001/4). Also see box 5 of the present study.

[13] Arabeyes, "The first open source project container for/by the Arab community". This information is no longer available online. However, further information concerning Arabeyes is available at: http://www.arabeyes.org/.

[14] Zeid Nasser, "The Penguin catches on in the Middle East: Factors that favour Linux", *The Star*, No. 103, 18-24 July 2002. Available at: http://star.arabia.com/article/0,5596,179_5119,00.html.

major components required to run Linux on the client side and is working on Arabic support for Motif, X-term and Mozilla.

The evolution of Linux and the benefits of Arabization are briefly reviewed in box 5.

Box 5. Evolution of Linux and the benefits of its Arabization project

The following series of events briefly describes the evolution of Linux:

(a) The open source movement was created in 1984 in the form of the GNU's Not Unix (GNU) Project;[a/]
(b) Linus Torvalds, a student from Finland released the first version (0.01) of Linux in 1991;
(c) Thousands of contributors continually refine/add/scrutinize the code;
(d) New applications are created/improved, thereby complementing the OS "kernel";
(e) Linux is touted as the most reliable and secure OS; and the most cost-effective enterprise/desktop solution;
(f) Towards the end of the 1990s, Linux held 27 per cent of the server market; Windows held 41 per cent.

The benefits of the Linux Arabization project include the following:

(a) Linux assures transparency by facilitating the availability of its code;
(b) Linux guarantees security;
(c) Linux enables the development/refinement of source (customization/enhancement);
(d) Linux enhances the technical abilities of the Arab world;
(e) Linux encourages a sense of self-reliance;
(f) Linux encourages experimentation;
(g) Linux provides business opportunities;
(h) Linux is extremely cost effective.

Source: Adapted from Arabeyes, "The first open source project container for/by the Arab community". This information is no longer available online. However, further information concerning Arabeyes is available at: http://www.arabeyes.org/.

a/ Available at: http://www.gnu.org/.

The popularity of Linux can also be attributed to the following factors:

(a) It is a free and legal option to pirated software. Rather than paying the relatively high price of licensing Microsoft Windows, users can utilize Linux, thereby making every aspect of computing cheaper. The growth of Linux at the server level will boost the adoption of desktops in companies. Furthermore, Arabized Linux at the desktop-level is going to be a serious alternative for Arab businesses in the future;

(b) It is more secure than other systems. This is based on the fact that: (a) users can fix problems easily; (b) Linux has not yet been effected by any computer viruses; and (c) UNIX OSs, from which Linux is derived, are more stable and scalable than Windows OSs.

Linux, which has the backing of a number of companies including IBM, Sun and Computer Associates, recently had a 5 per cent share of the business desktop market and has the potential to challenge Microsoft in the future. Indeed, the United Linux initiative, which aims to streamline operations, is likely to make Linux even more popular.[15]

In addition, Oracle is developing a low-cost, Intel-processor based computer that has no hard drive. The computer, which will be initially targeted at the education market, will use Netscape's Navigator and the

[15] Further information about Linux can be found on IBM, "The Linux line". Available at: https://www6.software.ibm.com/reg/linux/linuxline-r?S_ACTION=new.

Linux OS. It will carry 64 megabytes (MB) of main memory, and applications will be preloaded; users will upgrade the software using a CD-ROM drive and will also be able to boot up from the CD drive.[16]

1. *Functionality of Linux*[17]

UNIX sets out to provide users with a range of options and therefore provides a number of different versions and dialects. While this has led to a certain amount of competition among the various UNIX vendors, it has also paved the way for a better product.

Windows NT is developed by a single company, which determines the best features for the user. However, UNIX systems benefit from the input of various developers, which means that if one developer includes a specific functionality, other developers must also include it in the system or risk loosing customers. Linux takes this one step further by providing the source code for the OS and virtually all of the associated tools and programmes. Users can transmit information concerning a specific function that is missing in that they have direct access to the developer, who will include this function if he/she considers it to be significant; otherwise, the user can include it him/herself. In addition, Linux possesses dozens of tools that enable the system to be transformed to suit the needs of the users.

Another important consideration is configuration. Microsoft differs from Linux in that it determines the extent of configuration of any given system. The user has access to the internals of the Microsoft system through the registry; however, this is extremely cryptic and there is the danger that one small error will render the system unusable forcing a reinstall. With Linux, the fact that various copies of the kernel exist enables users to recover if the need arises.

While beginners find Linux more difficult to use than other systems, owing to the fact that it is accessible through long paths or obscure file names, it does have certain positive aspects. The Linux system is open, and therefore its configuration can be precisely tailored to the needs of the user. All versions are equipped with tools that allow the system to be expanded.

Linux has many advantages in terms of reliability and functionality. Users who are familiar with UNIX will not have too many problems with Linux; whereas knowledge of Windows NT is only relevant to a single product, which has been bought from a single vendor. Indeed, even knowledge and experience of one Microsoft system, for example, Windows 95, is of limited value with regard to Windows NT.

It is possible to conduct a comprehensive comparison of Linux and Windows OSs based on the following metrics: flavours; graphical user interface; text mode interface; cost; bugs; OS software installation; OS pre-installed; application software installation; obtaining application software; application software; viruses; software restrictions; playing on each others network; hardware devices supported by the OS; hardware that the OS operates on; clustering; security; multiple users; networking; hard disk partitions; swap files; file systems; file hierarchy; hidden files; case; modems; scripting; printer drivers; help; user data; and shutting down.[18]

Finally, while Windows NT and Linux both have their strengths and weaknesses with regard to hosting, it is worth bearing in mind that Windows NT hosting is more expensive than Linux hosting, and that the Linux OS tends to be more stable and reliable.[19]

---

[16] James Niccolai, "Oracle's Ellison promises $150, Linux-based computer", *IDG.net*, July 1999. Available at: http://www.infoworld.com/cgi-bin/displayStory.pl?990728.piellison.htm.

[17] This section has been adapted from *The Great Linux-vs-NT Debate*, "Linux: A matter of choice". Available at: http://www.linux-tutorial.info/Linux-NT_Debate/An_Editorial_on_LINUX.html.

[18] Michael Horowitz, "Linus vs Windows (a comparison)". Available at: http://www.michaelhorowitz.com/Linux.vs.Windows.html.

[19] Interspeed, "Linux hosting vs Windows NT hosting". Available at: http://www.interspeed.co.nz/linuxvswinnt.htm.

2. *Linux Arabization activities*

The objective of the Linux Arabization project, which was initiated in 1999, is to ensure complete Arabic support and the free distribution of Arabic Linux.[20]

Linux Arabization activities include a Saudi project, jointly funded by the Saudi Computer Society, King Abdulaziz City for Science and Technology, Atheer Internet Services and Compaq, which focuses on developing tools and applications for handling Arabic text and manipulating it in a Linux environment, and the provision of training for software developers.[21]

It is also worth noting that Linux4Arab.com offers complete Arabic support for Linux. In addition, Linuxvision.com has released Sheba, which provides complete Arabic support for Linux.

The features of Sheba include the following:

(a) "Supports most Gtk+ and GNOME applications for Arabic input and output;

(b) Supports text mode programs in Linux console and gnome-terminal for Arabic input and output;

(c) Supports a lot of other applications for Arabic output, like Netscape browser and KDE;

(d) Automatic configuration for some programs;

(e) The ability to change Sheba options for one or more programs without affecting the global options;

(f) Supports ISO 8859-6 and CP1259 charsets".[22]

The nature of Arabic language support for Linux and similar systems has seriously hampered the adoption of Linux in the Arab region. The majority of Arabic language support solutions are proprietary and do not fit in with the open source nature of Linux and UNIX-type systems. In this regard, Sheba, which is a proprietary solution, provides a degree of Arabic support to the GNOME user environment. The Linux User Group of Saudi Arabia, with the assistance of King AbulAziz City for Science and Technology, has been carrying out pioneering work in this area.[23]

While several volunteer groups have also attempted to launch initiatives related to Arabic support, lack of structure, organization and commitment has rendered the majority of these endeavours obsolete. However, Arabeyes, an open source non-profit volunteer grass-roots organization that was recently founded, has survived where others have fallen by the wayside.[24] Its activities include the incorporation of Arabic language support into existing open-source programmes, namely, VIM and PuTTY, and fixing certain Arabic issues with regard to applications that already have Arabic support, namely, Yudit, XFree86 and LyX.

Various projects have also been initiated by Arabeyes volunteers, including the following:

(a) *QaMoose*: an interactive English-to-Arabic dictionary;
(b) *Duali*: an Arabic spellchecker which is in its early stages;
(c) *Katoob*: an Arabic text editor;
(d) *Akka*: a system that provides Arabic support under the Linux/UNIX console;
(e) *Quran*: a browsable audible multilingual Holy *Quran*.

---

[20] *Linux Today*. Available at: http://linuxtoday.com/news_story.php3?ltsn=1999-08-06-002-05-NW.

[21] *Linux Today*. Available at: http://linuxtoday.com/news_story.php3?ltsn=2000-08-18-010-06-OS-CY-SW.

[22] *Linux Today*, "Linux4Arab.com: Complete Arabic support for Linux", July 2000. Available at: http://linuxtoday.com/news_story.php3?ltsn=2000-07-15-027-06-NW-CY-SW.

[23] Saudi Linux. Available at: http://www.linux.org.sa.

[24] Arabeyes.org. Available at: http://www.arabeyes.org.

In addition, Arabeyes is working on a translation project to provide an Arabic interface for well-known user environments (KDE and GNOME) and various programmes. Volunteers have also contributed to a documentation project that has facilitated the availability of HOWTOs and FAQs for Linux and Arabic support. Above all, Arabeyes is an umbrella group for Arabization projects in that it provides all the necessary tools and the means to preserve development and maintain continuity in this area.[25]

A brief explanation of Arabeyes is provided below.

(a) Arabeyes evolved into its present form on 29 July 2001;

(b) It was conceived by Arabs to address the lack of Arabic language support in Linux/UNIX platforms;

(c) Arabeyes offers the required know-how regarding all aspects of Arabization;

(d) It helps, supports and educates people on all issues regarding open source and Linux/UNIX;

(e) The group is a rallying-point for all those interested in advancing the cause of Arabs and computers;

(f) Arabeyes is a perpetual and unique project.[26]

Arabeyes, which has contributed to improving Arabic support in Linux, provides certain benefits for certain users. These are detailed below.

(a) The average user benefits from a productive system that makes it possible to access Arabized software;

(b) Small companies profit from a free and secure system that is financially sound, and does not entail the need for restrictive commercial licenses;

(c) Big corporations benefit from a cheaper alternative to proprietary closed-source commercial software, and also from competent programmers and developers with knowledge of issues related to Arabic software;

(d) Government offices and establishments receive the same benefits as big corporations and they also gain access to better security;

(e) The Arab programmer society acquires a better means of communicating, sharing and learning from one another;

(f) The Arab community benefits from being able to use free native Arabic software without license restrictions.[27]

Other Linux Arabization activities include the following OSS regional workshops:

(a) *Linux Arabization Workshop (Cairo 2-3 September 2001)*[28]

The Linux Arabization Workshop (LAW 2001) set out to do the following:

---

[25] "Arabic language support for Linux and other Linux-like systems has been a major barrier for Linux adoption in the Arab World", *The Star*, 8-14 August 2002, No. 106. Available at: http://star.arabia.com/article/0,5596,185_5358,00.html.

[26] Arabeyes, "The first open source project container for/by the Arab community". This information is no longer available online. However, further information concerning Arabeyes is available at: http://www.arabeyes.org/.

[27] The Arabeyes Manifesto. Available at: http://www.arabeyes.org/papers/manifesto/benefits.html.

[28] IBM, "Linux Arabization Workshop". Available at: http://www-5.ibm.com/eg/events/linuxagenda.html.

(a) Unite Linux experts and technical enthusiasts in the Arab world;
(b) Exchange ideas and experiences regarding Linux Arabization techniques;
(c) Coordinate efforts through the creation of mechanisms for cooperation and standardization.

Participants were requested to do the following:

(a) Encourage the Arab world to benefit from OSS in general and Linux in particular;

(b) Create a mechanism to foster the exchange of information and the coordination of efforts in the world of Arabization;

(c) Ensure proper standardization of Arabic Linux;

(b) *Open Source Software Workshop (Jordan, 15-16 December 2002)*[29]

A survey carried out during the workshop revealed more than half of the participants or their companies were familiar with and, in many cases, already using Linux, Mozilla, Sendmail, Apache, OpenOffice or other popular Open Source packages.

Finally, Linux user groups exist all over the world.[30] The Arab Linux User Group (ALUG)[31] is the first Linux user group for the Arabic language.[32] ALUG is a non-profit organization that is dedicated to promoting Linux in the Arab world by disseminating information on Linux, facilitating access to the system; and providing assistance during the initial stages of Linux use. The group includes engineers, corporate users, consultants, and journalists. Amongst its other activities, ALUG provides assistance to users by helping them to procure the latest Linux software and technical help.

D. ISSUES PERTAINING TO ARABIC DOMAIN NAMES[33]

Internet penetration in the Arab countries is illustrated in table 3. While it is estimated that 10 per cent of the 300-million strong Arabic world population speaks English, the number of Internet users in the region only amounts to some 3.5 million or some 1.3 per cent of the Arab population. This scenario emphasizes the importance of Arabizing contents, tools and applications and domain names. Domain names (see box 6) are utilized by *inter alia*, web sites and portals; e-mail addresses; file transfer protocols (FTP), Telnet and transmission control protocol/Internet protocols (TCP/IPs), namely, structured query language (SQL) (client/server); and Gopher.

Box 6. Arabic domain names

The Internet, which originally evolved in the United States, only supported 7-bit ASCII code (English characters); the domain name system (DNS) supports alphabets, numbers and hyphens in the ASCII code. However, the need to improve the global accessibility of the Internet has strengthened the importance of providing support for multilingual characters in domain names. As a result, expressing domain names in Arabic is becoming increasingly important and a number of solutions and implementations have been introduced, based on non-open standards.

Two non-profit international organizations that are attempting to promote multilingual issues and standards on the Internet are reviewed below:

[29] Information and Technology Association of Jordan (Intaj). Available at: http://www.intaj.net/calendar/event.cfm?id=232.

[30] LinuxWaves.com. Available at: http://www.linuxwaves.com/linuxlinks/Linux_User_Groups/.

[31] Available at: http://www.alug.linux4arab.com/.

[32] "First Arabic language Linux user group", *Linux Today*, 17 August 1999. Available at: http://linuxtoday.com/news_story.php3?ltsn=1999-08-17-002-05-NW-CY.

[33] A.M. Asiri, R.I. Al-Fayez and A.H. Al-Zoman, "Domain names: when is it going to be in Arabic?", November 2001. Available at: http://www.isu.net.sa/Library/domain_name_in_arabic.pdf.

Box 6 (*continued*)

(a) *Multilingual Internet Names Consortium*[a/]

The Multilingual Internet Names Consortium (MINC) is a non-profit, non-governmental, international organization that promotes (a) the multilingualization of Internet names, including Internet domain names and keywords; (b) the internationalization of Internet names standards and protocols, and (c) technical coordination with international bodies;

(b) *Arabic Internet Names Consortium*[b/]

The Arabic Internet Names Consortium (AINC) was founded in April 2001 in Amman to coordinate efforts amongst Arabic speaking communities, further the growth of online Arabic content and facilitate the development of new information technologies for the benefit of Arabic culture, language, education and development of skills, with specific attention to the Arabization of Internet names.

Source: Adapted from: Saudi Network Information Center (SaudiNIC), "Domain names in Arabic". Available at: http://www.saudinic.net.sa/arabicdomain/arabic_domains.htm.

a/ Available at: http://www.minc.org/.

b/ Available at: http://www.ainc.tn/.

TABLE 3. INTERNET PENETRATION IN THE ARAB COUNTRIES

Rank	Country	Number of subscribers	Number of users per account	Number of users	Percentage of Population
1	United Arab Emirates	220 000	3	660 000	24.44
2	Bahrain	35 000	3	105 000	16.67
3	Qatar	25 000	3	75 000	10.27
4	Kuwait	55 000	3	165 000	8.25
5	Lebanon	75 000	3.5	262 500	6.56
6	Jordan	35 000	6	210 000	4.57
7	Palestine	12 000	5	60 000	3.53
8	Oman	28 000	3	84 000	3.36
9	Tunisia	70 000	4	280 000	2.89
10	Saudi Arabia	190 000	3	570 000	2.59
11	Egypt	70 000	8	560 000	0.82
12	Morocco	55 000	4	220 000	0.73
13	Algeria	45 000	4	180 000	0.6
14	Libyan Arab Jamahiriya	4 000	5	20 000	0.4
15	Syrian Arab Republic	8 000	4	32 000	0.18
16	Yemen	3 500	4	14 000	0.08
17	Sudan	7 000	4	28 000	0.08
18	Iraq	500	25	12 500	0.06
	Total	938 000		3 538 000	1.29

Source: A.M. Asiri, R.I. Al-Fayez and A.H. Al-Zoman, "Domain names: when is it going to be in Arabic?", (November 2001). Available at: http://www.isu.net.sa/Library/domain_name_in_arabic.pdf.

The creation of Arabic domain names (ADNs) entails a need to address problems pertaining to linguistic issues, Arabic top-level domains (TLDs), technical solutions and root servers. Despite the fact that some companies provide ADNs, the following problems exist:

(a) Non-standard and incompatible solutions;
(b) Multi-registration;
(c) No recognition of ADNs by the Internet Corporation for Assigned Names and Numbers (ICANN);
(d) Isolation of Arabic networks.

Therefore, demand for ADNs has led to the establishment of a number of organizations including the Arab Internet Names Consortium (AINC), which includes an Arabic Linguistic Committee (ALC) that sets out linguistic guidelines.[34]

ALC aims to do the following:

(a) Define the accepted Arabic character set for ADNs;

(b) Define TLDs of the ADN tree structure, namely, Arabic generic TLDs (gTLDs) and country code TLDs (ccTLDs).

ADNs must preserve the unique features of the Arab world. They must also encourage Arabic language speakers to access Arabic sites. The results of a survey on Arabizing Internet domain names are illustrated in table 4 (Unicode is reviewed in annex I). These findings indicate that: (a) technologies must serve the Arabic language; (b) character folding must be not supported with separate character treatment; and (c) solutions must consider the unique properties of the Arabic language.

TABLE 4. RESULTS OF A SURVEY PERTAINING TO ARABIZING INTERNATIONAL DOMAIN NAMES

	Questions	Answers	Percentage	Recommendation
1	Using *Tashkeel*	Yes	17.25	Must not be allowed
		No	43.82	
		Use with no effect	36.13	
			2.80	
2	Using *Kasheedah*	Yes	28.61	Must not be allowed
		No	34.79	
		Use with no effect	30.36	
			6.24	
3	Allowing folding	Yes	35.01	Must not be allowed
		No	52.69	
			12.30	
4	Folding *waw* and *waw* with *hamza*	Yes	28.38	Must not be allowed
		No	50.00	
			21.63	
5	Folding *yeh* and *yeh* with *hamza*	Yes	30.62	Must not be allowed
		No	47.99	
			21.39	
6	Folding different states of *alef*	Yes	41.26	Must be allowed
		No	35.20	
			23.55	
7	Folding *heh* and *teh-marbotah*	Yes	29.20	Must not be allowed
		No	47.26	
			23.55	

[34] The ALC status report of April 2002 is available at: http://www.saudinic.net.sa/arabicdomain/status-report-april2202.doc (in Arabic and English).

TABLE 4 (*continued*)

	Questions	Answers	Percentage	Recommendation
8	Joining words	Merge	6.06	Spaces must be supported as part of a domain name
		Space	45.45	
		Dash	33.33	
		Other	3.03	
			12.13	
9	Numbers	Latin	27.27	Must be allowed to use both and they must be folded into one form
		Arabic	15.15	
		Both	45.45	
			12.13	
10	Arabic gTLD	Translation	41.90	Technologies must serve the language and the properties of language must be considered when choosing TLDs.
		New	43.94	
			14.16	
11	Mapping Arabic to English gTLD	One letter	15.81	Full Arabic words must be used to represent an English gTLD, namely, شركة for com, or شبكة for net
		Whole word	47.02	
		Root	15.16	
		New	14.42	
			7.58	
12	Arabic ccTLD	Standards	60.26	Names specified by the Arab Standards and Metrology Organization standard (two letters per country) must be used
		Anything	23.20	
		Other	1.93	
			14.63	
13	Criteria of choosing a gTLD	Standards	25.91	
		Number of letters	19.95	
		Clarity	23.61	
		Matching	22.19	
		Other	8.35	

Source: SaudiNIC. Available at: http://www.saudinic.net.sa/arabicdomain/survy-results.htm.

Note: Generic top-level domain.

Activities related to ADNs

(a) *Internationalization of DNSs*

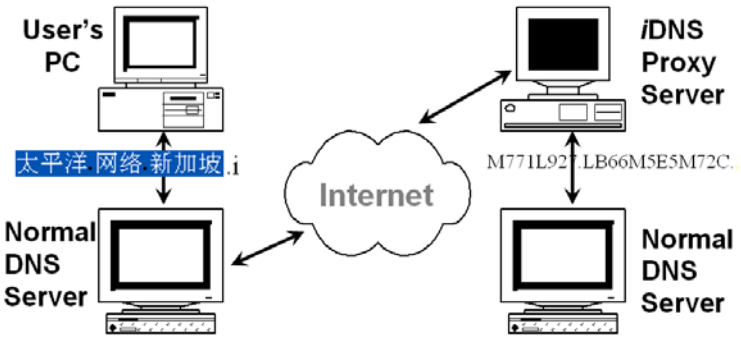
The internationalization of DNSs contributes to a multilingual Internet and provides an international solution that encompasses all languages and which can be tailored to the local needs of the native language. It also ensures backward-compatibility and inter-operability, thereby avoiding disruptions in the operation of the DNS (URL and e-mail).

Within this context, it is necessary to ensure the following issues are resolved:

(a) DNS must resolve the issue of multilingual characters;
(b) ADNs must be able to operate concurrently with English DNSs;
(c) ADNs must be compatible with DNS servers, clients and Internet protocols;
(d) ADNs must incorporate backward-compatibility;
(e) ADNs must be compatible with IPv4 and IPVv6 extensions;
(f) Consistent canonicalization must be ensured;
(g) Minimal configuration for end-users and changes on the client/server must also be ensured.

The processing of a non-Latin language, in this case Korean, is illustrated in figure VIII.

Figure VIII. DNS operations for an iDNS



Source: T.T. Wee, "*i*DNS; Internationalized domain name system". Available at: http://apng.net.edu.cn/idns/sgnic-dns.ppt.

(b) *i-DNS.net International*

i-DNS.net International announced in 2000 that it planned to provide key enabling multilingual technology to the Information and Decision Support Center (IDSC) of Egypt to enable the registration of domain names in Arabic within the ".eg" (Egypt) domain name space.[35]

IDSC governs the use of domain names ending with ".eg". The adoption of the internationalized domain name system (i-DNS) pioneered by i-DNS.net means that Egypt became the first country in the Middle East to endorse and implement country-level domain names in a local language. Therefore, anyone wishing to set up an .eg domain name can do so in Arabic. This also means that web sites that have English domain names can also have a corresponding one in Arabic, or any other local language.[36]

iDNS solutions for internationalizing DNS are reviewed in box 7.

Box 7. iDNS solutions for internationalizing DNSs

iDNS provides a server-based solution (iBind), a client-based solution (iClient), or a client + server combined solution (iResolve). iDNS complies with row-based ASCII compatible encoding (RACE), fully supports UTF-5, DUDE, AMC-ACE-Z, LACE, CIDNUC, SACE and other local encoding. iDNS standardization involves: (a) client encoding; (b) client-to-DNS server transport; (c) DNS zone file encoding (UTF-5 or UTF-8); and (d) compatibility with other protocols.

Source: I-DNS.net International, "Papers and presentations". Available at: http://www.i-dns.net/company/pressroom/papers.html.

E. Importance of Arabic ICT standards

ICT involves the specification, design and development of systems and tools that deal with operations that include the capture, representation, processing, security, transfer, interchange, presentation, management, organization, storage and retrieval of information. Information is either coded in written form using the character set of a language, or in a spoken form using the phonetic elements of a particular language. Multimedia systems use both written and spoken forms of language coding and will not perform the above-mentioned operations without the necessary standards. Each language has its particularities, and these must be taken into consideration with regard to standards. The Technical Committee 8, (TC-8) of the Arab

[35] Available at: http://www.i-dns.net/.

[36] i-DNS.net International, "News", 13 September 2000. Available at: http://www.i-dns.net/newsroom/news/I-EG000913-01.html.

Industrial Development and Mining Organization-Center of Standardization and Metrology (AIDMO-CSM) (see annex IV) drafted a standard for the use of Arabic language on the Internet in 2001.[37]

ICT standards that govern the following areas must be developed:

(a) Inputting information: keyboard layout, character-set and phonemes set;

(b) Processing information: compression standards of text and speech, natural language processing (NLP) algorithms and control codes;

(c) Transferring information over networks: Internet: standard transfer protocols and standard markup Languages;

(d) Output: displayed or printed character sets and page formatting;

(e) Application software: e-commerce, e-documentation and e-publishing;

(f) Terminology;

(g) Testing procedures to assure and control the quality of software;

(h) Issuing certificates pertaining to conformity to standards;

(i) Unifying procedures for accreditation/certification of testing;

(j) Quality control/quality assurance (QC/QA) of ICT hardware concerning electro magnetic compatibility, health requirements and user interfacing.

Standardization is an important element of the science and technology (S&T) agenda of all nations. Given that the ICT sector is experiencing rapid growth in the Arab world and is expected to form an essential component of its economy, the development of standards for Arabic in ICT is imperative.

The unique aspects of the ICT sector that are of particular relevance to the Arab world are detailed below:

(a) It is considered to be one of the largest sectors in the world with high value added potential, and a higher growth rate than the average growth rate of industry as a whole;

(b) Unlike other sectors, it has the ability to penetrate all aspects of daily life, namely, homes, schools, offices and factories;

(c) It incorporates the software industry, which has huge potential in the Arab world in terms of developing applications tailored to local needs.

Consequently, absence of regionally unified Arabic ICT standards might result in the following scenarios:

(a) ICT systems that cannot communicate between each other easily and efficiently;
(b) A number of fragmented ICT markets;
(c) Sub-quality software products.

In addition, those concerned with Arabic ICT standards can preserve the unique nature of the Arabic language by ensuring that the following actions are carried out:

[52] AIDMO, “Use of Arabic on the Internet”, a draft, (Tunis, 2001) (in Arabic).

(a) The swift production (or approval) of Arabic ICT standards by Arabic regional National Standarization Bodies (NSBs) that are affiliated to AIDMO-CSM;

(b) The formulation of Arabic ICT standards, which cannot be adequately adapted or translated from international standards, based on specifically relevant studies involving linguists, computer scientists, electronics engineers and standardization specialists;

(c) The enforcement of certain standards related to transparency, portability, efficiency, bilingualism and respect for the unique qualities of the Arabic language by NSBs in the IT field.

Arabic ICT standards have a vital role to play in e-commerce and online ventures in the Arab world in that they contribute to effective inter-changeability of data in Arabic among Internet users and they ensure conformity in terms of language use on the Web. According to the Arab Advisors Group (AAG), there is great potential for early stage e-commerce ventures in the region.[38] The Strategic Research Service of AAG, which is based on comprehensive primary research of eight countries, namely, Bahrain, Egypt, Jordan, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates, releases reports on the Internet and telecommunication markets in these countries. AAG bases its assessment of e-commerce business models on e-commerce pillars", which include Internet usage penetration levels, infrastructure development, education levels, financial services and income levels.[39]

In addition to e-commerce, other applications that have potential in the Arab world include, *inter alia*, e-government, e-banking and e-learning, all of which can be seriously facilitated by the development of comprehensive and unified relevant Arabic standards.

However, while standardization has the ability to promote Arabic e-commerce, e-government, e-content and e-learning activities, and therefore to contribute to economic and social transformations in the Arab region, a number of problems exist. Incompatible standards have long plagued the Arabic software arena and these must be unified and adhered to. Indeed, the Arabization of software is a process that encompasses a great deal more than the mere translation of the user interface from any given language into Arabic. In this context, the provision of localized Arabic software is more challenging than the provision of software in other languages (see box 8).

Box 8. Internationalization problems for the Arabic language

(a) *Character codeset and standard encoding*

There are more than 32 Arabic codesets in the PC environment or in vendor specific implementations (IBM, WANG, HP and UNISYS). Certain codesets are designated as 7-bit; others as 8-bit. The recommended codeset for representing Arabic script under the UNIX environment is the ISO 8859-6 (see chapter I, section F).

(b) *Character shaping*

The same Arabic character may correspond to up to four different glyph types. The glyph type of a character depends on the position of the character within a word.

(c) *Character fonts*

Arabic is always written in cursive form, whereby characters are linked together as if they were written by hand. The rules that govern linking are well defined; however, the font must be adapted to the cursive nature of the Arabic script, and the display device must be able to join all characters avoiding blank columns between characters.

[38] Arab Advisors Group. Available at: http://www.arabadvisors.com/.

[39] Michael Pastore, "Arab world holds potential for e-commerce", *CyberAtlas*, (January 2001). Available at: http://cyberatlas.internet.com/big_picture/geographics/print/0,,5911_568861,00.html.

Box 8 (*continued*)

(d) *Text direction*

One of the major problems with Arabic text is that Arabic is written from right to left, and mixed Arabic/Latin strings include text in both directions on the same line. In fact, text is stored in sequential order in the backing store. Logical or backing store order corresponds to the order in which text is typed. The conversion from backing store format to a readable format must be handled transparently by applications in low-level text rendering routines.

(e) *Global screen direction and mirror effect*

Based on the right-to-left writing direction of the Arabic language, the common way to read a document is from the top-right corner. This means that all user interface elements (menus, pick lists and edit boxes) must have a right-to-left orientation.

(f) *Numerals and Hindi digit shapes*

Numbers are read from left to right, despite the fact that Arabic text is read from right to left. In certain Arabic North African countries, digit glyphs are Arabic in form, namely, 1, 2 and 3. In some Middle Eastern countries, digit glyphs are in Hindi form. The display these digit representations must be user configurable.

(g) *Arabic vowels and collating sequences*

Arabic vowels have a specific status in Arabic text. They are represented by diacritics, which resemble Latin accents and are displayed above or below a consonant letter. From a collating sequence or a pattern search point of view, a word with vowels and the same word without vowels, has the same intrinsic value.

(h) *Neutral characters*

For technical reasons and to be able to display text correctly in right-to-left mode, some application screens or forms are initially built in a left to right direction. Some characters must be able to take the global writing direction despite their own direction value. Accordingly, it is necessary to define a set of neutral characters, which are able to use the global writing direction.

(i) *Dual keyboard management*

The use of Arabic characters (ISO 8859-6 includes both ASCII and Arabic characters) necessitates a dual keyboard management system.

Source: Adapted from F. Amara and F. Portaneri, "Arabization of graphical user interfaces", *Arabization of a user interface*. Available at: http://www.langbox.com/staff/arastub.html.

F. EXISTING STANDARDS

Standards pertaining to Arabic are reviewed in box 9. Security standards and specifications are highlighted in box 10 (also see chapter IV, section C). Standards for keys, hash functions, digital signatures, encryption and certificates are reviewed in chapter IV, section C, 3. Arabic and international organizations that deal with aspects pertaining to ICT Arabization and ICT standards related to Arabic language use in information society applications are briefly reviewed in annex IV.

Box 9. Standards pertaining to the Arabic language

(a) *BS 4280: 1969 (1983)*[a/]

This is a consistent system for the transliteration of Arabic characters. Arabic words in written form can be transcribed by replacing individual Arabic letters with corresponding letters from the Roman alphabet.

(b) *ISO 233: 1984*[b/]

This deals with the transliteration of Arabic characters into Latin characters. It follows the principles of stringent conversion to permit international information exchange and aims to provide a means for international communication of written messages in a form that permits their automatic transmission and reconstitution. It cancels and replaces ISO R 233-1961.

(c) *ISO 233-2: 1993*

This simplifies the stringent rules of ISO 233:1984 and facilitates the processing of bibliographic information, namely, catalogues, indices and citations.

(d) *ISO 233-3: 1999*

This is an upgraded version of ISO 233-2: 1993.

(e) *ISO 639: 1988*

This provides codes for the representation of languages, and represents Arabic as *ar*.

(f) *ISO 6438: 1983*

This contains a set of 60 African graphic characters and their coded representations for the interchange of bibliographic information.

(g) *ISO 8859-6:1987 (ASMO 449E)*

This is an 8-bit code closely related both to 7-bit ASCII and to ASMO 449; while the lower 128 positions are identical to ASCII (ISO 646), the upper 128 positions contain the Arabic characters of ASMO 449 in analogous places, in addition to extra graphic and control characters.[c/]

(h) *ISO 9036: 1987 (ASMO 449)*

This is an Arabic 7-bit coded character set for the interchange of information.

(i) *ISO/DIS 11822*

This is an extension of the Arabic alphabet coded character set for the interchange of bibliographic information.

(j) *ISO-10646 (Unicode)*

Unicode is a 16-bit code (see annex I).[d/]

---

a/ British Standards Online. Available at: http://bsonline.techindex.co.uk/.

b/ International Organization for Standardization, "Welcome to ISO online". Available at: http://www.iso.ch/.

c/ Further information concerning the ISO 8859 series can be found on the web site page "The ISO 8859 Alphabet Soup". Available at: http://czyborra.com/charsets/iso8859.html.

d/ Further information concerning Unicode can be found at the Unicode Home Page. Available at: http://unicode.org/; and Stonehand.com. Available at: http://www.stonehand.com/.

Box 10. Standards pertaining to information technology security

The standards listed below are of relevance to IT security.

(a) *ISO/IEC TR 13335*: Guidelines for the management of information technology (IT) security;
(b) *ISO/IEC 15408: 1999*: Evaluation criteria for IT security;
(c) *BS 7799: 1999*: British standard for information security management;
(d) *Common criteria (CC) for IT security evaluation*: Provides an evaluation of IT security systems;
(e) *Common data security architecture (CDSA) Version 2.0*: Secures inter-system communications;
(f) *FIPS Pub 140-1*: Security of United States federal information processing standard for security requirements for cryptographic modules.

Source: Diffuse, "Information security standards". Available at: http://www.diffuse.org/secure.html.

II. CLASSIFICATION OF ARABIC LANGUAGE STANDARDIZATION ISSUES

A. TRANSLATION OF WEB PAGES

Given that more than 400 million people worldwide (see table 1) have access to the Internet in a language other than English, the importance of translating existing standards, particularly those pertaining to ICT, into Arabic, cannot be underestimated. The ability to translate web sites from English to Arabic has a direct effect on the amount of Arabic content on the Web, and is a means of increasing the number of Arabic language online users.

Several efforts have been exerted towards this end, including the translation into Arabic of 249 ISO standards that were adopted at the Seventeenth Meeting of the Higher Consultative Committee for Standards of AIDMO, which took place in May 2001.[40] However, further endeavours must be carried out with regard to the translation of existing standards pertaining to ICT systems and protocols into Arabic. While many translation tools support translation across language pairs, very few tools include Arabic as an option.

A selection of translation tools are reviewed below; only Babylon and Ajeeb support Arabic:

(a) *SYSTRANet*[41]

This service translates personal files (TXT, RTF and HTML formats), web pages and plain text. Users select the most suitable language translation pair and the custom topic dictionary. Translations are available in a few seconds;

(b) *WordReference Personal*[42]

This tool has the ability to translate any word on any web page in coordination with *Collins Dictionaries*. It operates with Windows 95/98/NT and is available in two forms: one that is compatible with Internet Explorer, and the other with Netscape;

(c) *AltaVista Babel Fish*[43]

Powered by SYSTRAN translation engines, Babel Fish is a speedy service that can translate chunks of 150 words of text copy into and from English from a large set of languages;

(d) *FreeTranslation.com*[44]

This is an easy-to-use site that provides rapid translations. It also offers a web translator that can translate a uniform resource locator (URL) into one of the several supported languages;

(e) *WordLingo*[45]

This site enables users to send e-mails, which are automatically translated into the language of choice upon delivery. Attachments can also be translated;

---

[40] Arifonet.org. Available at: http://www.arifonet.org.ma/aidmo/databases/standard/liste_doc_standard_traduits_Iso.htm (in Arabic).

[41] Available at: http://www.systranet.com/systran/net.

[42] Available at: http://www.wordreference.com/index.htm.

[43] Available at: http://world.altavista.com/.

[44] Available at: http://www.freetranslation.com/.

[45] WorldLingo. Available at: http://www.worldlingo.com/products_services/email_translation.html.

(f) *LogoMedia.net*[46]

LogoMedia.net provides access to a wide range of free translation services, including the translation of web pages and e-mails, document upload and the use of machine translation software and direct text machine translation;

(g) *Babylon*[47]

This tool instantly provides relevant information, translations and conversions of any word or value that a user clicks on;

(h) *Ajeeb*[48]

This service has the ability to translate Arabic and English and can translate both text and web pages.

B. INTERNET

Issues and sites related to the use of Arabic on the Internet are reviewed in the following sections.

1. *Character set*

The dominance of the English language in the sphere of ICT is diminishing. Indeed, ensuring that normal, non-specialized computers are able to operate in non-English languages is becoming easier and cheaper. While it is possible to use all languages on the Internet, this is not always a perfect process. For example, messages sent in non-Latin character sets do not always arrive in the correct form. This situation may change in the future; in the meantime, awareness of the manner in which language operates in the context of the Internet will benefit those utilizing non-Latin scripts (see box 11).[49]

Box 11. Computers and characters

There are two main reasons why the use of non-Latin scripts poses problems with regard to the Internet: (a) computer systems, including Disk Operating System (DOS), Windows, Apple Macintosh (Macintosh) and UNIX, differ in the manner in which they handle non-English; and (b) the restrictions of the network itself, which affects non-Latin scripts or languages that require more characters than the unaccented A-Z.

So why does it matter that Windows and Macs are different? Because users don't always know what kind of computer is at the other end of the cable, which can cause confusion. Computers transmit each letter of a text as a numeric value, and the correspondence between the letters and numeric values are known as character sets. Computer types such as Macs and PCs are all completely standardized with regard to their character set for English and those characters have identical numeric values on all computers. However, everything else is basically a free for all. Not only do the character sets for non-European scripts such as Arabic or Chinese differ depending on the computer type, so do the accents and characters for some European languages, namely, the Spanish ñ, German ß and Danish ø. While the numeric value of e is the same on a Mac as on a PC, the numeric value of é is quite different.

In this context, any software that handles relations between PCs and Macs must take this difference into account. Normal programmes and word processors usually handle this; unfortunately, network software is not as well behaved. If you put an é into an email message from a PC, the software will send the numeric value of the é under the PC system. The Mac will receive this character value, and display the corresponding character in its own character set, which is not an é, thereby leading to a confused or corrupted message. This situation is even trickier with regard to Arabic.

[46] LogoMedia.net. Available at: http://www.logomedia.net/.

[47] Babylon. Available at: http://www.babylon.com.

[48] Ajeeb. Available at: http://tarjim.ajeeb.com/ajeeb/default.asp?lang=1.

[49] Non-European Computing, "European and non-European languages on the Net; a survey of problems and solutions". Available at: http://www.hf.uib.no/smi/ksv/char.html.

Box 11 (*continued*)

This scenario makes communication between different computer systems difficult. However, even Mac-to-Mac or Windows-to-Windows communication can be impractical, when dealing with French, Russian or Chinese. This is because of the restrictions on the Web itself. For historical reasons, the Internet has a formal rule that certain types of messages, in particular e-mail, can only use a limited number of characters. While any Mac or PC has the ability to display approximately 220 different characters in any font, only some 90 of these can be used on the Web; this is in line with the 7-bit rule, which basically blocks the use of characters not present in English in e-mail and other software.

Source: Adapted from Non-European Computing, "European and non-European languages on the Net; a survey of problems and solutions". Available at: http://www.hf.uib.no/smi/ksv/char.html.

2. *Arabic on the Web*

The ways in which Arabic can be represented on the Internet are briefly reviewed below.

(a) *Images*

Given the problems of reading non-Latin scripts on the Internet, web pages in Arabic are often converted into pictures. This means that a full page of original text is saved and presented as an image. While this is slower to download than a regular page, special tools do not have to be used to view the page and Arabic fonts do not have to be installed. Web images are often saved in the GIF format (see box 17).

(b) *Acrobat or PDF files*

This process is similar to the one mentioned above, except that images saved in this format are technically superior to GIF images in that they allow the use of the zoom function. This method does not require the installation of Arabic fonts and pages can be viewed on any system. *Al-Hayat*, an Arabic daily newspaper based in London is available on the Web in Acrobat and can be accessed through Acrobat Reader.

(c) *Text*

The browser must be equipped with the relevant Arabic fonts to display Arabic as text. Unlike the two previous options, this ensures full Web functionality. Text files are much smaller than images and therefore, can be downloaded more rapidly. However, as indicated above, there are several systems for writing Arabic, which means that if the provider uses one system and the browser reads another, an erroneous output will ensue. Basically, there are three different ways of displaying Arabic:

(i) *ISO Arabic (ISO-8859-6)*

Macintosh uses ISO Arabic, and a few extra characters;

(ii) *Windows Arabic (CP-1256)*

This is probably the most commonly used Arabic text format on the Internet. The alphabet in this format is the same as ISO up to the Arabic letter of *dad*; however, the latter letters are different.

(iii) *Unicode (UTF-8 and UTF-16)*

Unicode includes all possible scripts; it is utilized by many web pages and can be supported by some browsers.

Several browsers support Arabic, including the Mozilla derivation of Netscape and the independent iCab browser, which are detailed below.

(a) *Mozilla*[50]

This is an offshoot of Netscape that resembles Netscape 6. It supports all Arabic character sets, namely, MacArabic, ISO, Windows Arabic and Unicode and it selects an Arabic font automatically. Mozilla is a huge and demanding application that requires 20 to 28 MB of memory; in contrast iCAB requires 3 to 4 MB and Explorer needs 4 to 10 MB. Mozilla would seem to be one of the best browsers for Arabic;

(b) *iCab*[51]

This is a new, fast Web browser that has been developed for Macintosh. While this was not designed for Arabic, it does take that language into consideration and can display MacArabic, ISO Arabic, Arabic Windows and Unicode, with no user settings, or visits to a menu; Arabic simply appears on the page.

3. *E-mail*

Certain e-mail software, namely, Standard Eudora version 2, supports Arabic almost perfectly, with some minor glitches. Arabic can be sent and received correctly by installing the Arabic EudoraTables file.[52] Eudora must be set up to display the message in an Arabic font. Alignment can vary depending on the system being used, namely, whether users are dealing with Macintosh or Windows. Macintosh systems have separate Arabic period and colon characters. While Windows does not have this function, it does have a common English and Arabic period. Therefore, any period, colon or similar punctuation in an Arabic text from Windows (or UNIX) will cause the line to break in two, before and after the period. The text before the period will be displayed to the left, and the text after the period, to the right, as in English text.

4. *Usenet News*[53]

The Usenet News site is essentially a worldwide bulletin board that facilitates the free exchange of information on any subject and gives all users an equal opportunity to participate in discussions. Usenet News divides subject areas into newsgroups, also known as conferences, forums and discussion groups, thereby directing people to a particular area of interest. Each newsgroup relates to one subject or topic. Some newsgroups deal with very specific topics and other newsgroups are more general in nature. The use of Arabic has so far been rather limited on this site.

5. *WHOIS*[54]

WHOIS data, which includes information associated with domain names, is used for a variety of purposes, including identifying and verifying online merchants, locating information on investigations carried out by consumer protection and other law enforcement authorities, determining the availability of a domain name for registration, enforcing intellectual property rights, addressing cyber-attacks and resolving technical network issues. The services of WHOIS have been available on the Internet since the early 1980s, and are widely used. According to an online survey of more than 3000 participants from businesses, Governments, Internet service providers (ISPs), registrars, individuals and non-commercial organizations conducted in 2001 by the ICANN Domain Name Supporting Organization, Internet users generally indicated that WHOIS data was important and they supported measures to improve its accuracy.[55]

---

[50] Mozilla.org. Available at: www.mozilla.org.

[51] iCab. Available at: www.iCab.de.

[52] Non-European Computing, "Using Eudora for languages other than English: the Eudora tables emporium". Available at: http://www.hf-fak.uib.no/smi/files/eudtab.html.

[53] Time McLellan, "An introduction to Usenet News" (1997). Available at: http://www.islandnet.com/~tmc/html/articles/usentnws.htm.

[54] Available at: http://www.whois.net/.

[55] Domain Name Supporting Organization, *Draft Final Report of the Name's Council's WHOIS Task Force*. Available at: http://www.dnso.org/dnso/notes/whoisTF/20020625.TFwhois-report.htm.

C. LANGUAGES[56]

The Web is a widely used platform for information systems. Many companies or information providers have linked their databases to the Web to ensure the worldwide accessibility of their information and to benefit from web browsers. However, problems exist with regard to presenting home pages that are not written in Latin scripts. It is not possible for all users to install fonts for all character sets. Well-known web browsers tend to support ISO-8895-1 and several western languages. Web browser providers have implemented local solutions to meet national and local needs. However, such browsers are usually limited to a few languages, which means that Arabic web browsers cannot display documents written in Greek.

Search engines will search for the word or words that have been typed in by a user, sometimes using fuzzy logic for stemming and query expansion. However, users cannot search for documents written for example, in Arabic, if their terminal does not support Arabic input. This can make it difficult to search for the names of organizations and buildings, which are only available on the Internet in the local language. Moreover, while many records are accessible via the Internet, it is difficult to read records from a foreign terminal if they are encoded in non-ASCII codes.

Some tools and applications are based on Unicode, which appears to resolve character set and data exchange problems. Unicode is a single 16-bit code, which enables the encoding of more than 65,000 characters; this means that it covers most languages. Most OSs, namely, Microsoft, IBM, DEC, Sun and Macintosh use Unicode. HTML 4.0 supports Unicode as the reference character set for web pages. Tango, from Alis Technologies supports all business languages for display and input purposes. Multilingual Mosaic of Accent Technologies is based on Unicode. The Microsoft Front Page Editor and Java also support Unicode. Database systems such as Oracle, Sybase, Informix, and Adabas provide Unicode support.

Internationalization and multilingual text access rely on character sets, user interfaces and the Web, to ensure correct data representation, interpretation, manipulation and presentation. A number of working groups have been established to deal with these issues. For example, several groups focus on the internationalization of HTML, URL and HTTP within the framework of the W3C World Wide Web Consortium.[57] Other American and European initiatives examine multilingual information access and meta data.[58]

The internationalization of the Web is highlighted in box 12.

Box 12. Internationalization of the Web

"(a) *Hypertext markup language*

Internationalization of HTML means that it should be able to deal with non-Western characters in the text, such as Arabic, Chinese, Thai, etc. The non-Western characters should be represented in a HTML document properly. Further, it should support their display and other operations correctly, since a HTML document will contain text fragments in multiple language. The former HTML versions were based on ISO Latin 1, which support only West-European languages. Special characters for French, German, and Spanish languages (ISO-8859-1/Latin 1) are supported. Thus, rendering special characters like ç, Ü or ñ can be performed easily and do not require any special tools. For instance, the French word "*français*" is encoded as "français" and all web browsers can render it correctly. How other ISO Latin 1 special characters are defined, can be found at the Center for the Study of Languages site.[a/]

[56] Hachim Haddouti, "Multiliguality issues in digital libraries". Available at: http://haddouti.de/euromed.html.

[57] W3C World Wide Web Consortium. Available at: http://www.w3c.org.

[58] Multilingual Information Access Working Group. Available at: http://www.cs.columbia.edu/~klavans/Activities/MIA/home.html.

Box 12 (*continued*)

Later, options have been added to support other European languages (e.g. ISO- Latin 2), but the real problem of internationalization remains open. HTML 4.0 and future versions will be designed to support the Universal Standard Character Set, Unicode … and bi-directional multilingual text.

"(b) *Extensible markup language*

EXtensible markup language (XML) 1.0 is a relatively new markup language in the Web. It becomes easy to define document types, which can be globally shared on the Web. According to user defined rules (e.g. those of document types) XML documents can be parsed and validated. The separation between a document content, structure and layout will benefit the interoperability of applications highly. XML is based on ISO 10646/Unicode.

XML requires that any XML processor accept both UTF-8 and UTF-16 (Unicode encodings). It also supports the attribute (xml:lang), which indicates the language of the contents.

"(c) *Uniform resource locators/identifiers*

URLs are Web resource addresses that are limited to ASCII. However, an address is the facility and "guide" to find someone or something. This restriction to ASCII forces other people (e.g. Arabic) to provide their addresses in ASCII, even if these addresses are for local needs only.

In the framework of the internationalization of WWW … the use of UTF-8 seems to be the preferred character encoding for uniform resource identifiers (URIs). UTF-8 is an encoding of Unicode into 8-bit characters. This encoding is suitable because on the one hand it allows representing URL or URIs in the local character code (Unicode-based), and on the other hand it is compatible to the current approach, i.e. the current URLs based on ASCII can still be used.

"(d) *Hypertext transfer protocol*

HTTP is an 8-bit protocol … that uses language tags within Accept-Language and Content-Language (RFC1766). The same language codes are used for HTML and XML. However, a few compatibility issues, such as MIME type while transmitting Unicode text, should be addressed. Furthermore, the HTTP protocol ignores in some implementations the charset parameter (e.g. Charset = ISO8859-1 or charset = US-ASCII, which leads to incompatibility problems for users. In general, HTTP seems to be the least challenging problem of the internationalization of WWW.

"(e) *Multilingual-HTML*

This solution, developed at the University of Library and Information Science (ULIS) in Tsukuba Japan,[b/] aims at presenting multilingual documents, even if a browser and a client platform do not contain and support the required fonts. This system has been extended to allow the input in multiple languages. Using a Multilingual-HTML (MHTML) Server, users can display and search multilingual documents from any Java-enabled browser."

Source: Hachim Haddouti, "Multilingual text access", *Survey: Multilingual text retrieval and access*, February 1999, pp. 14-16. Available at: http://haddouti.de/review-issue.doc.

a/ Rice University, "Center for the Study of Languages". Available at: http://langcenter.rice.edu/.

b/ "Multilingual-HTML Browser Project". Available at: http://mhtml.ulis.ac.jp/.

D. HARDWARE[59]

The issue of standardization pertaining to computer hardware must be addressed, particularly with regard to unifying standards for displays, keyboards and other computer peripherals.

[59] F. Amara and F. Portaneri, "Arabization of graphical user interfaces", *Arabization of a user interface*. Available at: http://www.langbox.com/staff/arastub.html.

1. *Displays*

There is a lack of standardization in the way that Arabic characters appear on displays. This has been exacerbated by the fact that Arabic text is written from right to left and numbers are written from left to write; mixed Arabic/Latin strings include text in both directions on the same line (see (d) and (f), box 8).

2. *Dual keyboard management*

A dual keyboard management system allows the user to switch from one language to another with a single keystroke (see (i), box 8). This solution must either include an engraved keyboard, which incorporates both ASCII and Arabic letters on each key, or more flexibly, a set of keyboard stickers that can be installed on the existing keyboard.

3. *Optical character recognition*

The methods and algorithms for optical character recognition (OCR) for Latin scripts are difficult to use in relation to Arabic owing to the fact that Arabic uses connecting letters. The main problem in this regard lies in the ability to extract a single letter from a word. This issue is reviewed in greater detail in chapter III, section G.

4. *Peripherals*

While many peripherals, including printers, plotters and head mounted displays (HMD) lack Arabic language support (through a lack of unified standards), some are fully bilingual, namely, Alis A3356 and A3156 printers, which provide a solution for heavy-duty printing of forms using Arabic and Latin scripts and which include transparent built-in Arabization.[60] Alis also produces the al-khattat series, which configures any of the 4000, 4050, 5000 and 8100 HP laser jet printer to print Arabic text. It operates in a text mode environment under UNIX or Windows. Standard host application software can send bilingual Arabic and Latin text to print easily and quickly without special Arabizing host software or device drivers.

E. APPLICATIONS

Many applications lack the requisite standards for Arabic language utilization, including those related to word processing, spreadsheets, search engines, chatting, cryptology, OSs and other software.

1. *Word processing*

ArabTeX extends the capabilities of TeX/LaTeX, and can generate Arabic writing from an ASCII transliteration. It comprises a TeX macro package and an Arabic font in several sizes, only available in *Naskhi*. ArabTeX is compatible with Plain TEX and LaTeX. Limited support is available for other languages that are customarily written in the Arabic script. ArabTeX can be used in e-mails, and texts in certain Arabic standard encodings can also be processed.[61]

A selection of Arabic word processors is listed below:

(a) Al Kateb (DOS);
(b) El-Qalam (Windows);
(c) Bazar 2.0 (DOS);
(d) InText (DOS): supports many languages including Arabic;
(e) The Universal Word (Windows 95 and NT);
(f) ArabTeX (Unix and DOS);
(g) Nisus (Macintosh): WorldScript II support;

[60] Globalis International. Available at: http://www.alis.com/arabe/Planet_printing_solutions.html.

[61] ArabTeX. Available at: http://user.cs.tu-berlin.de/~ishaq/arabic/asm/node26.html.

(h) ArabFont (DOS); collection of Arabic fonts;

(i) WinText 2.7.2; French programme for Arabic that also handles other non-Latin scripts;

(j) WorldWrite 3.0 (Macintosh and PowerPC); multilingual word processor that can handle any language supported by the Apple language kit.

2. *Search engines*

Despite the fact that web pages in the Arabic script constitute a very small portion of the Internet, tools must be developed to enable non-HTML based databases and textual data in Arabic scripts, to be indexed, searched and published. The basic architecture of a search engine that is capable of indexing, searching and displaying results from the Web is examined in a study that emphasizes different techniques used for presenting Arabic text in non-Arabic environment systems.[62] The proposed engine provides two methods for locating records in a bibliographical database. The first method enables the user to browse through predefined categories, namely, subject or alphabetical listings. The second method permits free text searching against the fielded database content. Results are displayed using a hit list web page, and the entire record is displayed on the browser screen upon selection of an entry.

3. *Operating systems*

The DOS market created most of the Arabic implementation standards, with some reinforcement from Microsoft Arabic DOS and Arabic Microsoft Windows.[63] A user moving from these systems to a UNIX environment will find the same Arabic features and uses. Some PC solutions cannot be implemented in the UNIX environment. For example, in Arabic DOS character applications, users press Right Shift+Left Shift to toggle their keyboard layout between Arabic and Latin mode. If users press the same key sequence on an alphanumeric terminal, such as VT100 or Wyse, the UNIX OS and therefore the application will receive nothing from the serial line connection. Hence, it will be unable to use this sequence to switch between two internal logical keyboard mappings. Arabic OSs include the following: (a) *Mussa'ed Al Arabi*; (b) OS/2 WARP; (c) Bilingual Microcomputer System; (d) arabic.com; (e) LANGBOX and XLANGBOX.

[62] A. Al-Kharashi, "A Web search engine for indexing, searching and publishing Arabic bibliographic databases". Available at: http://www.isoc.org/isoc/conferences/inet/99/proceedings/posters/085/.

[63] F. Amara and F. Portaneri, "Arabization of graphical user interfaces", *Arabization of a user interface*. Available at: http://www.langbox.com/staff/arastub.html.

III. PRIORITY ISSUES IN ARABIC LANGUAGE STANDARDIZATION

Priority issues pertaining to the harmonization of ICT standards related to Arabic language use in information society applications are reviewed in this chapter.

A. ARABIC CHARACTER SETS AND THEIR STANDARDS[64]

The existence of multiple Arabic character sets poses a major problem for the use of Arabic on the Internet. The character set enables conversions between textual and digital forms. The latter form is necessary for transmission along the physical layer of any network.

A brief history of Arabic character sets is reviewed below:

(a) 1981: CUDAR-U was the first standard Arabic character set, using 7 bits per character;

(b) 1982: The Arab Standards and Metrology Organization (ASMO) produced AMSO-449 (7 bits), which became the basis for all subsequent standards and plays a similar role to that of ASCII for Latin characters;

(c) 1986: ASMO standard 708 (8 bits) became the international standard ISO-8859-6. Since then, it has gained widespread acceptance and has been particularly used in the Arabized Macintosh system.

By the late 1980s, more than 20 Arabic character sets had been developed. With the spread of PCs and the Microsoft Windows operating system in the 1990s, the Microsoft Windows Arabic code page (MSCP-1256) became the utilized standard (see annex II). MSCP-1256 permits the simultaneous use of Arabic and French in addition to other display control characters.

Unicode provides a unique number for every character regardless of platform, programme and language. It was jointly developed by the International Organization for Standardization (ISO-10646) and the Unicode Consortium. Prior to the existence of Unicode, there were hundreds of different encoding systems for assigning these numbers. For example, the European Union requires several different encodings to cover all its languages. Unicode addresses the problem of the plurality of character sets and as a result, it has been promoted on the Internet as the character set of choice in new protocols. An important characteristic of Unicode is that it uses 32 bits, which encodes over 35,000 characters, to represent all characters in use across the globe. The Arabic UCS encoding corresponds to the ASMO-449 code set. UCS has various transfer encoding syntaxes such as: UCS-4 (32-bit), UCS-2 (16-bit), UTF-16 (multiple 16-bit), UTF-8 (multiple 8-bit), and UTF-7 (multiple 7-bit). UCS has detailed character property tables and algorithms (for example bi-directional text display), which are particularly suited for Arabic. Moreover, it provides characters for text directionality. Further information on Unicode, including its Arabic character code tables is included in annex I. The Microsoft Arabic code table and additional Arabic code tables are provided in annex II.

B. ARABIC DISPLAY ISSUES

Difficulties pertaining to the display of Arabic characters include the representation of Arabic text and the directionality of display (see box 13).

Box 13. Character set definition

(a) A coded character set (CCS) is a mapping from a set of abstract characters to a set of integers, namely, ISO-8859-6 (see chapter I, section F, box 9);

(b) A character encoding scheme is a mapping from a CCS to a set of octets, such as UTF-8;

[64] Sections A, B and C of this chapter have been adapted from B. Al-Badr, "Using the Internet in Arabic: problems and solutions," Internet Summit 1998 (INET'98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

Box 13 (*continued*)

(c) A transfer encoding syntax is a transformation that is applied to character data that enables it to be transmitted, namely, base64 encoding. It is used to transform encoded text into a format that can be transmitted through the use of specific protocols.

The knowledge of the character set used is necessary for the correct encoding at the transmitting end, and to decode text at the receiving end. Specifying the parameters of a textual transmission requires both (a) a set of labels for specifying the character set (which can be done in the MIME headers), encoding scheme and transferring syntax used; and (b) a procedure for attaching these labels to data. The labels are typically registered with the Internet Assigned Numbers Authority.

Source: Adapted from B. Al-Badr, "Using the Internet in Arabic: problems and solutions," Internet Summit 1998 (INET'98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

Box 14. Representing Arabic text and displaying bi-directional text

(a) *Representing Arabic text*

"When discussing the display of characters, it is important to distinguish between the characters themselves and their visual representation, called glyphs. While a character is a letter, a series of characters are visually represented as a series of glyphs. This is particularly important in Arabic, where shapes of characters depend on the context. To display Arabic text correctly, a context analysis program is needed to select the right shape of a character (glyph) depending on the context. The context is not necessarily the preceding and following characters only. Arabic script is highly decorative, and many ligatures (a glyph for multiple characters) are used, especially in stylized fonts. This implies that Internet client programmes that display Arabic (such as web browsers) must employ contextual analysis or rely on an underlying operating system to do that. Finally, Arabic has a number of diacritic marks that are written above and below the characters to aid in pronunciation (see figure V)".

(b) *Displaying bi-directional text*

"Before Arabic text is displayed, "it must be reordered correctly on the screen. This is an important issue because most computer systems are designed to display text left-to-right, and also because bi-directional text must be simultaneously displayed on the same text line (e.g., Arabic words and numbers).

"Unicode defines a direction property for each character and provides directionality algorithm for the display of bi-directional text. The directional property of Arabic characters is strong right-to-left, while the characters of other languages are strong left-to-right.

"The text directionality algorithm uses a set of directional ordering codes to influence the ordering of text. These codes are used for embedding one language into another and for overriding the default direction of text".

Source: B. Al-Badr, "Using the Internet in Arabic: problems and solutions", Internet Summit 1998 (INET'98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

C. ARABIC E-MAIL

The specification of e-mail on the Internet can be divided into the following two major components:

(a) Mail transport, which is managed by an application-level protocol that runs over TCP, namely, SMTP or the newer extended SMTP (ESMTP);

(b) E-mail message format, which is specified in RFC 822 and is updated by the MIME standard.

The standards pertaining to the format of the transport and messages hamper the exchange of Arabic e-mail messages.[65] The SMTP standard specifies the transfer of ASCII-text messages, which indicates that non-ASCII characters may not reach their destination. The RFC 822 message format standard identifies two main components: the message header and the message body. The latter is composed of lines, each of up to 1,000 7-bit ASCII characters.

Therefore, the following issues must be resolved with regard to Arabic e-mail:

(a) The correct transportation of messages that are 8-bit encoded;

(b) The specification of the language and character set used in a particular message on the basis that transporting e-mail does not involve a prior exchange of information about content, as in HTTP.

It is possible to solve these problems through the MIME standard, which allows for the labelling and structuring of message contents using RFC 822 headers by introducing a new set of header fields that are added to the message header. This permits transmitting binaries and non-ASCII text through e-mail by encoding them in ASCII. The user can also specify the character set used in a message.

MIME issues are highlighted in box 15.

Box 15. Encoding an e-mail message body and representing body and header character sets

(a) *Encoding e-mail message body*

"Using MIME, 8-bit content in the message body can be encoded using 7 bits. The transfer encoding syntax is specified in the header field "Content-Transfer-Encoding", which can take on the values: 7-bit, 8-bit, binary or base64. Base64 is a transfer encoding syntax that represents groups of 24-input bits as output strings of four encoded characters. The encoded characters are from an alphabet of 64 ASCII characters".

(b) *Representing body and header character sets*

"The content of a MIME message body is labelled using the special header field: "Content-Type", which has, as a parameter, the character set specification field "charset". … Using MIME, the message header can contain non-ASCII text by using inline labelling".

An example of a non-ASCII header is: Subject: =?ISO-8859-6?B?SWYgeW91IGNhbiB=?= ,

where the encoding "B" refers to base64.

"The names of character sets that are used in MIME headers must be registered with IANA.[66] The registered character set names for Arabic include: ISO-8859-6 (ASMO-708), ISO-9036 (ASMO-449), Windows-1256 and ISO-10646 (Unicode)".

Source: B. Al-Badr, "Using the Internet in Arabic: problems and solutions", Internet Summit 1998 (INET'98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

D. ARABIC WEB BROWSING[67]

The two main components of specifications related to the Web are detailed below:

[65] B. Al-Badr, "Standards for supporting Arabic on the Internet", *Proceedings of the First King Saud University Workshop on Internet Arabization*, Riyadh, 18 May 1997 (in Arabic).

[66] The Character Sets Registry. Available at: http://www.iana.org/assignments/character-sets.

[67] Sections D and E of this chapter have been adapted from B. Al-Badr, "Standards for supporting Arabic on the Internet", *Proceedings of the First King Saud University Workshop on Internet Arabization*, Riyadh, 18 May 1997 (in Arabic).

(a) HTTP is a page transfer protocol that permits the transport of Arabic pages in 8-bit character sets;
(b) HTML is a page description language.

The major issues surrounding the use of Arabic on the Web are the labelling of the relevant character set and marking up Arabic pages in HTML.

The internationalization of the HTML standard has prompted the introduction of new features that facilitate the use of Arabic on the Web (see box 12). These features are now incorporated in HTML 4.0, which is based on Unicode. The internationalization features that are relevant to Arabic include: (a) character set definition and language tagging; (b) bi-directional text markup; and (c) controlling cursive joining behaviour (see box 16).

Box 16. Relevant internationalization features for Arabic

(a) *Character set representation and language tagging*

Character sets can be defined in the following three ways:

(a) On the "charset" attribute of the "A HREF" element;
(b) Through the "META" element in a HTML document header with the MIME-like content-type header;
(c) Through the HTTP header sent ahead of the document MIME tags.

Language tagging, which is different from character set specification, facilitates high-level operations such as searching, sorting, hyphenation and spell checking. Specifying the language of a text block is achieved through the "lang" attribute, which can be part of most HTML elements, including the "span element".

(b) *Bi-directional text markup*

HTML bi-directional specifications promote the use of the Unicode directional text display facilities, and stipulate that if the web client (browser) claims to display bi-directional text, then it must use the Unicode algorithm. Text directionality is encoded in the directional property of the characters. However, additional directional markup is necessary for direction-neutral text and tables. HTML offers higher-level markup constructs to control text direction, which are similar to the direction characters function of Unicode.

Neutral characters are an example of bi-directional text that might require additional markup. These type of characters determine the position of a double quote when its sits between an Arabic and a Latin letter. In that case, two marks are defined: the left-to-right mark "&lrm" and the right-to-left mark "&rlm", which are invisible characters with no other effect. The direction attribute "dir" indicates the base directionality of the text and can take either "LTR" or "RTL". DIR is attached to block-type elements, namely, <HTML>, <P>, <LI>, and <TD>, and also sets the default value of the "ALIGN" attribute. Furthermore, it affects the correct placement for bullets and facilitates the formation of bilingual tables.

(c) *Cursive joining behaviour*

HTML offers the zero-width joiner "&zwj" and the zero-width non-joiner "&zwnj" to mark up unusual cases of cursive text. HTML 4.0 specification define these joiners as follows:

"The zwnj entity is used to block joining behavior in contexts where joining will occur but shouldn't. The zwj entity does the opposite; it forces joining when it wouldn't occur but should. For example, the Arabic letter "HEH" is used to abbreviate "Hijri", the name of the Islamic calendar system. Since the isolated form of "HEH" looks like the digit "five" as employed in Arabic script (based on Indic digits), in order to prevent confusing "HEH" as a final digit five in a year, the initial form of "HEH" is used. There is no following context (i.e., a joining letter), however, to which the "HEH" can join. The zwj character provides that context".

Source: Adapted from B. Al-Badr, "Using the Internet in Arabic: problems and solutions", Internet Summit 1998 (INET'98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

The confusion regarding choice of character set, and the complexity of displaying Arabic text, have restricted the growth of Arabic on the Internet. Accordingly, web publishers have resorted to other techniques to overcome such problems (see box 17).

Box 17. Other Arabization techniques

(a) *Text as image techniques*

It is possible to convert pages of text into images, which can be displayed by most web browsers. However, problems associated with this solution include, most notably, the huge increase in page size, which means slower page download and display. Furthermore, it is not possible to do text operations such as search, selection, copying, or editing on the text images, unless the image undergoes character recognition, which is not necessarily accurate. In addition, the publisher of the pages must deal with the issue of increased storage space and complex publishing procedures (also see chapter II, section B, 2).

(b) *Proxy conversion*

In this case, the web server deduces the characteristics of the client (browser), particularly the supported character set, and then supplies a version of the page that is suitable for the client. The pages can be stored in Unicode and converted to the required character set.

(c) *Java applet*

This Java-based solution relies on intelligence to display Arabic in a Java applet and enables it to manage the display of Arabic text in the browser.

(d) *Automatic detection*

Browsers, including Sindbad from Sakhr, favour this technique, which utilizes intelligence to infer the character set of a Web page by analyzing its contents. The browser then switches to the inferred character set and displays the page. This is a heuristic approach that appears to be relatively accurate, owing to the fact that there are only two significant major Arabic character sets. Similar technology is used in the Alta Vista search engine to deduce the character set, which is capable of inferring character sets from a set of 25 languages (excluding Arabic).[a/]

In conclusion, Arabized browsers must have the ability to select a language and character set, load a font, accept Arabic input and display bi-directional text. Various browsers provide such capabilities (see chapter I, section B).

Source: Adapted from B. Al-Badr, "Using the Internet in Arabic: problems and solutions," Internet Summit 1998 (INET'98), Geneva, Switzerland, 21-24 July 1998. Available at: http://www.cetp.ipsl.fr/~porteneu/inet98/5f/5f_1.htm.

a/ Altavista. Available at: http://www.altavista.digital.com/.

E. ARABIC SEARCHING AND INDEXING

Searching and indexing tools are crucial for locating and organizing Internet information. These processes are more complicated for Arabic text than for English. Arabic is a language where words are derived from a root. Therefore, searching and indexing Arabic text relies on the root of a word. Moreover, the same word can have more than 100 combinations of prefixes and suffixes, which in English might precede words such as "with" and "for". Accordingly, search tools for Arabic must employ morphological analysis, which is an involved process and has its limitations. The large number of synonyms of Arabic words intensifies this problem. Furthermore, Arabic has a large number of combined word expressions. To search for an expression, one needs to use a logical operator such as "and" or "near". Current usage of Arabic includes many foreign words, which can be written in different ways, thereby leading to retrieval difficulties.

Several Arabic searching and indexing tools are available on the Internet. While Arabic is not officially supported, some of the standard search engines, namely, Alta Vista and Infoseek allow Arabic keywords to be entered to search for Arabic documents.[68] Arabic search engines include Ayna,[69] Naseej[70] and Alidrisi.[71]

F. SPEECH TECHNOLOGY

English to Arabic automatic translation are produced by: (a) App-Tek;[72] and (b) SYSTRAN, which incorporates English-to-Arabic software on IBM mainframes.[73]

Selected definitions pertaining to speech recognition are detailed below:[74]

(a) *Speaker adaptive system*

This modifies its operation to suit the characteristics of new speakers;

(b) *Speaker independent system*

This is the most demanding and expensive type of system to develop, and it is not very accurate. This system is designed to function for speakers of all types;

(c) *Speaker dependent system*

The speaker dependent system works in coordination with a specific speaker and is usually cheaper and easier to develop, though not as flexible, as the previous two systems;

(d) *Continuous speech system*

This is used for uninterrupted speech (connected speech) and therefore, it can be difficult to handle because it is difficult to find the start and end points of words;

(e) *Isolated-word system*

This is the simplest form of speech recognition and is used for interrupted speech (one word at a time);

(f) *Automatic speech recognition*

Automatic speech recognition (ASR) involves mapping an acoustic speech signal to text;

(g) *Automatic speech understanding*

This involves mapping an acoustic speech signal to some form of abstract meaning.

---

[68] GO.com. Available at: http://infoseek.go.com/.

[69] Ayna. Available at: http://www.ayna.com/.

[70] Available at: http://www.naseej.com/index2.html (in Arabic).

[71] Available at: http://www.alidrisi.com/.

[72] Available at: http://www.apptek.com.au.

[73] Available at: http://www.systransoft.com/.

[74] Comp.speech Frequently Asked Questions Web page, "What is speech recognition?". Available at: http://www.speech.cs.cmu.edu/comp.speech/Section6/Q6.1.html.

Automated systems that interact more effectively with human users by means of verbal and written communication will be more popular in the future and will greatly enhance human productivity. Previous research on ASR has mainly concentrated on Asian and European languages; other language groups, such as Arabic, have been investigated to a lesser extent. The Arabic language present problems for standard ASR systems owing to its complex morphological structure and the fact that written representation does not contain most of the vowels that exist in the spoken form, which makes it difficult to utilize textual training data.

The main root of failure of ASR in systems dedicated to standard Arabic can be attributed to particularities of the Arabic language, namely, geminate and emphatic consonants and vowel duration.[75] Work on automatic recognition of Arabic phonetic macro-classes and complex phonemes by multi-layer sub-neural-networks (SNNs) and knowledge-based system has been carried out. In addition, novel speech recognition models for Arabic were explored at a workshop held at Johns Hopkins University (Baltimore, 15 July – 23 August, 2002).[76] Speech recognition involves a wide variety of technologies including the following:[77]

(a) *Speaker-dependent speech recognition*

Speaker-dependent (SD) recognition can be used in cases where user-specific or language-specific vocabularies are required. Each recognition word is trained by the user to create voice templates. SD recognition has a 99 per cent accuracy rate;

(b) *Speaker-independent speech recognition*

Speaker-independent (SI) requires no end-user training and is designed for a specific language. SI technology is has an accuracy rate of more than 97 per cent;

(c) *Speaker verification speech recognition*

Speaker verification (SV) is similar to SD recognition, with the additional feature of being able to identify the original speaker;

(d) *Speaker adaptive speech recognition*

Speaker adaptive (SA) recognition can be used in single-user applications to update voice templates, thereby increasing accuracy over time by responding to changes in the environment and the voice of the speaker;

(e) *Dual recognition technology*

This integrates SI and SD recognition to improve accuracy;

(f) *Continuous listening*

Continuous listening (CL) enables products to respond to specific, discrete commands without pressing a button or waiting for a prompt. It offers both SD and SI recognition;

(g) *Word spotting*

This is an advanced version of CL and has the ability to extract key-words from a normal conversation. This technology endeavours to improve the human-to-machine interaction by creating a more natural language interface;

---

[75] J. Caelen and S.A. Selouani, "Regognition of Arabic phonetic features using neural networks and knowledge-based system: a comparative study", *World Scientific*. Available at: http://www.worldscinet.com/ijait/08/0801/S0218213099000063.html.

[76] Available at: http://www.clsp.jhu.edu/ws2002/groups/arabic/.

[77] Sensory. Available at: http://www.sensoryinc.com/.

(h) *Fast digits*

This is an optimized algorithm for rapid-digit string entry. It is compatible with voice dialling applications, namely, handsets and mobile phones.

Arabic-based technology applications are gaining importance and many projects related to Arabic speech recognition have been undertaken, including the joint venture that involves the King Abdulaziz City for Science and Technology in Saudi Arabia and a project to enable users to interact with computers over the telephone in Arabic. Arabic speech recognition products are available from IBM and the Linguistic Data Consortium (LDC).[78] The IBM ViaVoice speech recognition products for Windows, Macintosh and Linux, namely, the Arabic ViaVoice Millennium Edition are installed on Arabic Windows NT and Arabic Windows 2000.[79]

With regard to speech synthesis, a text-to-speech (TTS) system renders documents as spoken output. A text document provided as input to the TTS system can be produced automatically by human authoring or through a combination of methods. The speech synthesis markup language defines the form of the document.

A TTS system undertakes the following major processing steps to convert marked-up text input into automatically generated voice output:[80]

(a) XML Parse: an XML parser extracts the document tree and content from the incoming text document;

(b) Structure analysis influences the way a document is read;

(c) Text normalization performs an automated conversion of the written form into the spoken form;

(d) Text-to-phoneme conversion converts a set of words into a string of phonemes;[81]

(e) Prosody analysis ensures that features of speech, including pitch, timing, pausing, speaking rate and the emphasis on words, sound human;

(f) Waveform production utilizes phonemes and prosodic information.

A specification for a speech synthesis markup language for use in systems that need to produce computer-generated speech output, namely, voice and web browsers has been produced by W3C and is available on the Internet (see footnote 109). "A survey of existing methods and tools for developing and evaluation of speech synthesis and commercial speech synthesis systems" is also available on the Web.[82] In addition, exhaustive lists and demonstrations of speech synthesis systems can be located on various sites.[83]

Examples of commercial Arabic speech synthesis systems include: (a) *BeSTspeech*;[84] (b) *MBROLA Project*; [85] (c) *Sakhr Software* (also see chapter I, section B).

---

[78] Linguistic Data Consortium. Available at: http://www.ldc.upenn.edu/Catalog/search.html.

[79] IBM, "Arabic Via Voice Millennium Edition". Available at: http://www-3.ibm.com/software/speech/ar/.

[80] W3C, "Speech synthesis markup language specification for the speech interface framework", a working draft (8 August 2000). Available at: http://www.w3.org/TR/2000/WD-speech-synthesis-20000808.

[81] Phonemes are basic units of sound in a language. Each language has a specific phoneme set. In Arabic this conversion is vague since the same written word may have several spoken forms.

[82] DISC. Available at: http://www.disc2.dk/tools/SGsurvey.html.

[83] For example, Carnegie Mellon University, "List of software/hardware/information". Available at: http://www.speech.cs.cmu.edu/comp.speech/FAQ.Packages.html.

[84] "BeSTspeech from Berkeley Speech Technologies". Available at: http://www.speech.cs.cmu.edu/comp.speech/Section5/Synth/bestspeech.html.

[85] Available at: http://tcts.fpms.ac.be/synthesis/.

G. Optical character recognition

Arabic OCR programmes are much slower than comparable programmes for Latin OCR. Research must be carried out in the area of NLP to improve Arabic OCR and to resolve the challenges related to the cursive nature of the Arabic script, character overlapping, variety of character shapes and diacritics.

OCR programmes for Arabic that were on the market as of 1994 include the following:

(a) TextPert 3.7 Arabic: This is produced by CTA, Inc., and runs on the Macintosh Arabic system;

(b) Al-Qari' al-Ali (Arabic Automatic Reader): This is produced by al-Alamiah Software Co., and runs on al-Nawafidh al-Arabiya, the Arabization programme for Windows;

(c) ICRA 4.0: This is produced by Arab Scientific Software and Engineering Technologies and is an application for Windows, with Arabic support.

TextPert 3.7 is a user-friendly programme, that produces acceptable results when used on high quality, simple texts; however, with regard to texts that utilize more complicated fonts, the programme recognizes very little.

Al-Qari' al-Ali is based on an algorithm that combines vector and bit-map analysis. Its first upgraded version offers a limited variety of means of controlling recognition performance. Results can be controlled by a spell checker, which despite limitations is surprisingly good, particularly for controlling words that have run together. The main drawback of this OCR programme is the considerable amount of time it takes to train for new fonts, particularly hand set fonts with many ligatures.[86]

OCR Arabic products include the following:

(a) Arabic OCR/World Processing Bundle;[87]
(b) Arabic Automatic Reader (Professional Edition);[88]
(c) OCR Office (Sakhr);
(d) OCR Professional (Sakhr).

The main features of OCR Office include the following:[89]

(a) High accuracy and a recognition speed of up to 500 characters per second;

(b) Flexibility; includes or excludes components of the image files and assigns separation points to set the frames sequences with ease;

(c) New formats for saving files; saves output files in HTML and Unicode formats;

(d) E-mail guide; sends text output files through e-mail programmes;

(e) Image auto rotation; adjusts image in the event of a position error on the scanner;

---

[86] The appraisals of these programmes were based on J.N. Bell and P. Zemanek, "Test of two Arabic OCR programs", *Arabic and Computers*. Available at: http://www.hf.uib.no/smi/ksv/arbocr.html. ICRA 4.0 was not tested by the authors owing to limited time and means.

[87] WorldLanguage.com, "Arabic OCR products". Available at: http://www.worldlanguage.com/Products/Arabic/OCR/Page1.htm.

[88] Ibid.

[89] Aramedia. Available at: http://www.aramedia.com/ocroffice.htm.

(f) Recognizes text only inside rectangles using the text box feature;

(g) Recognizes text in ill-formed tables;

(h) Uses non-rectangular shapes to set text size precisely inside any frame with any shape, excluding rectangles;

(i) Supports a large number of scanners;

(j) Bundling personal archiving programme; edits and saves images in files and folders.

OCR Professional, which is similar to OCR Office, has the ability to do the following:

(a) Open multiple image files simultaneously;
(b) Recognize coloured images (in Arabic and European scripts);
(c) Recognize Arabic/French images;
(d) Support the learning mode;
(e) Support four types of batch jobs;
(f) Correct, edit and spell check the resultant text whilst a batch job is taking place;
(g) Support more than 16 European languages;
(h) Incorporate 26 Arabic font libraries.

IV. LEGISLATIVE AND REGULATORY ISSUES FOR INFORMATION AND KNOWLEDGE MANAGEMENT

The production of most goods and services depends on ICT systems. Such systems are either utilized in the production process itself or as a component of the system that make the goods or services available to consumers. Furthermore, the relationship between ICT, telecommunication and power supply systems highlights the existence of increased interdependencies across corporate and sector boundaries, whereby all players are dependent on the ICT systems of one another.

Information and knowledge management (IKM) involves the recognition and development of general principles concerning data security. Basic concepts and principles relating to IKM must be developed and correctly utilized in the context of the legal system for regulatory and law-drafting purposes. Furthermore, the manner in which the legal order is described as a systematic entity must be revised. Legal regulation is a traditional means of managing risks and threats including issues related to data processing and ICT data security norms. The flow of information and information itself, have become increasingly vulnerable and at the same time, dependence on data processing and ICT is also on the rise. One regulatory challenge pertains to the anonymity of legal transactions resulting from the development of ICT. The traditional categories and classifications of transactions and the corresponding regulatory regimes do not have any bearing on the new technical reality.[90]

A. ICT TERMINOLOGY

ICT terminologies in all ICT-related Arabic standards must be unified. The Institute of Electrical and Electronic Engineers Standards Association (IEEE-SA), an international membership organization that has a complete portfolio of programmes for standards, provides an example of how such unification might be achieved.[91] The IEEE, which is the largest technical professional society in the world, coordinates standards projects. Coordination refers to distributing the project draft for review and comment, and ensuring that other organizations parties, and IEEE societies with a direct interest in the project are given an adequate opportunity to participate in the development process.

Every project must coordinate with the Standards Coordinating Committee 10 (SCC10).[92] SCC10, which is the sponsor of IEEE Standard 100 and *The IEEE Dictionary of Electrical and Electronics Terms*, is responsible for overseeing the use and development of terminology in IEEE standards. SCC10 screens terms and definitions to ensure the appropriate and consistent use of terminology. Accordingly, a similar committee must be established to unify ICT terminologies in all ICT-related Arabic language standards.

B. STANDARDS FOR THE UTILIZATION OF ARABIC LANGUAGE ON THE INTERNET

An Internet Best Practices Committee for the Arabic language must be formed to define a similar Arabic standard to the IEEE Standard 2001-1999, which defines recommended practices for web page engineering and Intranet/Extranet applications.[93] This standard addresses the need to effectively develop and manage web projects based on HTML, XML or other languages. This standard encompasses the issues of life cycle planning, identifying the audience, client environment, objectives and metrics, and includes recommendations on server considerations and specific web page content. Such a standard is intended to reduce site-management costs, reduce legal risks, facilitate user satisfaction and increase the productivity of web applications for both maintainers and users.

---

[90] Ministry of Finance of Finland, "Data security and law", in coordination with the Institute of Legal Informatics and the University of Lapland (1997). Availble at: http://www.urova.fi/home/oiffi/julkaisut/datasec.htm.

[91] Institute of Electrical and Electronic Engineers (IEEE), *IEEE Standards Association Home Page*. Available at: http://standards.ieee.org/.

[92] IEEE, "SCC10 coordination and the standards process", *IEEE Standards Association Home Page*. Available at: http://standards.ieee.org/reading/ieee/SB/Jul95/scc10.html.

[93] IEEE, "IEEE Standards Association approves recommended practice for Internet practices", *IEEE Standards Association Home Page*. Available at: http://standards.ieee.org/announcements/2001.html.

One organization that might make some contribution to this committee, which would also be concerned with intellectual property issues, is the Arab Society for Intellectual Property (ASIP) is a recognized intellectual property society, which develops and promotes intellectual property protection in the Arab world through its activities, seminars, courses and participation in international intellectual property organizations and bodies.[94] Intellectual property has become one of the major issues in modern economic policy in that wealth is increasingly tied to intellectual rather than physical assets. The dawn of the knowledge economy has arrived, where the real currency and key to wealth is knowledge. Trademarks and copyrights in relation to e-material such as music, software and artistic or literary creations are particularly important issues in the present e-world.

C. SECURITY ISSUES

Wireless Internet users will account for an estimated 48 per cent of all Internet users in 2005.[95] Therefore, security of data transmitted via the Internet is becoming an increasingly demanding issue. Basic data security qualities are described below:[96]

(a) Confidentiality: this is the disclosure of data to authorized users to protect their right to exclusive access;

(b) Integrity: this is the reliability, validity and authenticity of accessed data in relation to the original data;

(c) Availability: this is real time accessibility to data in terms of collection and processing.

1. *Vulnerability of ICTs*

Contemporary economies and societies are becoming completely dependent on ICTs owing to the fact that they have the ability to spawn opportunities, economic growth and new methods of propagating information. Indeed, ICTs have become such an integral component of the systems and utilities that are integral to daily life that malfunctions can cause disruptions in power supplies, telecommunications, health, payment systems, industrial sectors and any other system that utilize ICTs. Accordingly, several countries are now implementing national measures to safeguard systems that depend upon ICTs. *Critical Foundations – Protecting America's Infrastructures*, a report published in 1997 in the United States, considers American infrastructures to be extremely vulnerable. Sweden and Switzerland have conducted similar studies, and in 2000, the State Secretary Committee for ICT in Norway delivered a report on the status of work being carried out on ICT vulnerability in the country.[97]

Moreover, technology has rendered society more vulnerable to a wider range of threats than ever before, including the actions of individuals and those of nation States. A hacker can access computer systems that control critical infrastructures and which are linked to the Internet from anywhere in the world, thereby disrupting their normal functions with potentially serious consequences. The increasing complexity of ICT systems also represents a serious threat in that simple random technical errors, or the external influences of nature or individuals, can initiate unpredictable chain reactions.

Over the past decade, breaches of security have occurred in banks, hospitals and companies. While there is no direct link between the vulnerability of a system and the vulnerability of the technology used in the system, the increasing complexity of ICT systems and infrastructures entails the possibility that complicated malfunctions might occur with increasing frequency in the future.

---

[94] Arab Society for Intellectual Property (ASIP). Available at: http://www.aspip.org/ (in Arabic) and http://www.aspip.org/english/index.htm (in English).

[95] Computer Industry Almanac. Available at: http://www.c-i-a.com/pr_info.htm.

[96] Ministry of Finance of Finland, "Data security and law", in coordination with the Institute of Legal Informatics and the University of Lapland (1997). Available at: http://www.urova.fi/home/oiffi/julkaisut/datasec.htm. This analyses data security norms, the drafting of laws and case law in a number of countries.

[97] Ministry of Trade and Industry of Norway, "Society's vulnerability due to its ICT-dependence", Oslo, October 2000. Available at: http://odin.dep.no/nhd/norsk/enorge/024101-990018/index-dok000-b-n-a.html.

Therefore, it is necessary for modern society to implement measures to reduce vulnerability. Efforts must also be exerted to ensure that ICT infrastructures are robust enough to allow important functions to continue operating at all times. This goal can be achieved through a comprehensive strategy that highlights cooperation, exchange of information and flexibility in terms of ICT vulnerability (see box 18).

Box 18. ICT vulnerability

The public and private sectors face common challenges and are interdependent with regard to the issue of ICT vulnerability. The following elements are possible components of a strategy to reduce ICT vulnerability:

(a) *Partnership between the private and public sectors*

This will promote mutual trust and exchange of information.

(b) *Exchange of information*

The exchange of information among various types of organizations, in the private and public sectors, will promote greater awareness of ICT systems, which have become so complex that it is not possible for a single body to have a full overview of them.

(c) *Greater warning capability*

Implementing measures to share technical surveillance, notification and handling of security incidents among organizations will ensure that preparations are taken to stave off new threat scenarios.

(d) *Education*

ICT security education will improve knowledge of ICT vulnerability and security, and ensure that professional knowledge is updated.

(e) *Research and development*

More research and development (R&D) on vulnerability analysis will improve awareness of the issue of ICT vulnerability;

(f) *Security of infrastructures of critical importance*

Defining security requirements for infrastructures, namely, telecommunications and power systems, which are critical to other ICT systems, will improve security.

(g) *Adaptation of laws and regulations*

Adapting the legal framework to suit the dynamics of ICTs while balancing individual freedom and the protection of society will ensure that the rights of all people and organizations are maintained.

The following recommendations outline methods for realizing the above-mentioned strategy:

(a) *Centre for Information Assurance*

A centre must be established to coordinate efforts to improve ICT security and to contribute to more robust ICT infrastructure.

(b) *Increased R&D*

A strategic research programme must be established in the field of ICT vulnerability and security, in coordination with business and industry.

(c) *Education*

ICT training must be improved, focusing on security, at all levels. Moreover, specialized expertise in this area can be achieved by establishing ICT security as a separate subject at universities.

(d) *Risk and vulnerability analysis*

Efforts must be exerted to ensure that companies that depend on ICT systems use risk and vulnerability analysis;

Box 18 (*continued*)

(e) *User-targeted measures*

End-users must be prepared to implement and use security infrastructures, namely, electronic signatures;

(f) *Combating crime*

Minimum and maximum penalties for infringements of the law in this area must be adapted to the seriousness of the crime. In addition, given that ICT systems are not circumscribed by national boundaries, legislation must be applicable across the Arab world. The issue of personal privacy and the regulation of the relationship between technology and the individual must also be taken into account.

Source: Ministry of Trade and Industry of Norway, "Society's vulnerability due to its ICT-dependence", Oslo, October 2000. Available at: http://odin.dep.no/nhd/norsk/enorge/024101-990018/index-dok000-b-n-a.html.

Standards and specifications related to the above-mentioned issues are outlined in box 19, (a) below.

2. *Evaluation criteria for ICT security systems*

Universal security evaluation criteria are a means of measuring the capability of ICT security products. Relevant standards and specifications are outlined in chapter I, section F, box 10.

3. *Elements of ICT security systems*

The main elements of ICT security are reviewed in box 19.

Box 19. ICT security elements

(a) *Keys*

These are numeric values, which when applied to a message through the use of an encryption algorithm produce the encrypted version of a message. Each key belongs to one user. Standards and specifications pertaining to this area include the following:

(i) *RFC 1422*: Privacy enhancement for e-mail;
(ii) *RFC 2408*: Internet security association and key management protocol;
(iii) *RFC 2409*: Internet key exchange;
(iv) *RFC 2528*: Representation of key exchange algorithm keys in Internet X.509 public key infrastructure certificates;
(v) *RFC 2631*: Diffie-Hellman key agreement method;
(vi) *ISO/IEC 11770*: IT security techniques; key management;
(vii) *PKCS#8*: Private-key information syntax standard.

(b) *Hash functions*

These are used for generating and verifying digital signatures. Hash functions enable the software for creating digital signatures to operate on smaller amounts of data. Standards and specifications pertaining to this area include the following:

(i) *PKCS* #1: Rivest-Shamir-Adelman (RSA) signature based on ISO/IEC 9796-2;
(ii) *PKCS #5*: Password based encryption;
(iii) *RFC 1319*: MD2 message-digest algorithm;
(iv) *RFC 1320*: MD4 message-digest algorithm;
(v) *RFC 1321*: MD5 message-digest algorithm;
(vi) *Secure hash system/secure hash algorithm*: Conversion of large files into a condensed message digest;
(vii) *ISO/IEC 10118*: IT security techniques.

Box 19 (*continued*)

(c) *Digital signature*

This is a digitally signed hash result of a message, to which it is typically attached. A digital signature is unique to its message and permits the verification of the sender. Standards and specifications pertaining to this area include the following:

(i) *ISO/IEC 9796*: Digital signature scheme giving message recovery;
(ii) *CEN ENV 12388*: Medical informatics; RSA algorithm for digital signature services;
(iii) *ISO/IEC 14888*: Digital signature mechanism with appendix for messages of arbitrary length;
(iv) *Digital signature standard/digital signature algorithm*: Encryption of digital signatures;
(v) *PKCS #1*: RSA signature.

(d) *Encryption*

This is associated either with the encryption of the message or the generation of a unique digital signature based on the message. Standards and specifications pertaining to this area include the following:

(i) *FIPS 46-3*:[a/] Latest version of the data encryption standard (DES);

(ii) *FIPS 74*:[b/] Guidelines for using DES;

(iii) *FIPS 81*: DES modes of operation (under revision);

(iv) *ISO 9735: 1999*: Security of United Nations Electronic Data Interchange for Administration, Commerce and Transport (UN/EDIFACT) messages;

(v) *ISO/IEC 10181: 1996* Security for open systems interconnection;

(vi) *RFCs 1968 and 1969:* Point-to-point protocol encryption.

(e) *Certificates*

These are generally issued by certification authorities (CA). A certificate is an electronic record that binds a key to a particular holder. Standards pertaining to this area include the following:

(i) *ANSI X9.57*: Certificate management for financial services;
(ii) *FIPS 102*: Guidelines for computer security certification and accreditation;
(iii) *ISO/IEC 9594-8: 1995*: Open systems interconnection and authentication framework;
(iv) *RFC 2459*: Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile;
(v) *RFC 2510*": Internet X.509 public key infrastructure certificate management protocols;
(vi) *RFC* 2511: Internet X.509 certificate request message format;
(vii) *RFC 2527*: Internet X.509 public key infrastructure certificate policy and certification practices framework;
(viii) *ISO/IEC 14888*: Digital signatures with appendix*;*
(ix) *ISO 15782*: Management of security within banking IT systems.

Source: Diffuse, "Guide to information security". Available at: http://www.diffuse.org/secguide.html.

a/ United States Department of Commerce/National Institute of Standards and Technology (NIST), "Data encryption standard (DES)", *Federal Information Processing Standards Publication*, October 1999. Available at: http://csrc.nist.gov/ publications/fips/fips46-3/fips46-3.pdf.

b/ NIST, "FIPS". Available at: http://csrc.nist.gov/publications/fips/.

D. DIGITAL PRESERVATION STRATEGIES

While digital technologies enable information to be created, manipulated, disseminated, located and stored with increasing ease, preserving access to this information poses a significant challenge.[98] Digital preservation strategies include the following:[99]

(a) Refreshing: this is the copying of information without changing it and subsequently, its storage on newer media before the old media becomes inaccessible;

(b) Migration: this is the movement of digital information from one computer platform to another;

(c) Emulation: this is the preservation of the functionality and integrity of digital objects;

(d) Encapsulation: this groups together a digital object and any tools necessary to provide access to that object;

(e) Adherence to standards, namely, ISO Archiving Standards: this preserves access to digital information, by facilitating the transfer of information across computer platforms as technologies evolve.[100]

Resources that are encoded using open standards have a greater chance of remaining accessible after an extended period than those encoded with proprietary standards.

There are a number of standards for the various aspects of storing and accessing digital information, including the following:[101]

(a) Interoperability standards;
(b) Resource encoding standards;
(c) Resource identification standards;
(d) Resource description standards;
(e) Data archiving standards;
(f) Records management standards.

---

[98] Library of Congress, *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, (Washington D.C., Council on Library and Information Resources and Library of Congress, April 2002). Available at http://www.clir.org/pubs/reports/pub106/contents.html and K.H. Lee and others, "The state of the art and practice in digital preservation", *Journal of Research of the National Institute of Standards and Technology*, Volume 107, Number 1, January – February 2002. Available at: http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf.

[99] National Library of Australia, *Preserving access to digital information (PADI)*. Available at: http://www.nla.gov.au/padi/topics/18.html.

[100] *ISO Archving Standards – Overview*. Available at: http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html.

[101] National Library of Australia, "Standards", *PADI*. Available at: http://www.nla.gov.au/padi/topics/43.html.

V. CONCLUSION AND RECOMMENDATIONS

This study highlights Arabic language calligraphy and script; the Arabization of the Internet and relevant standards; and the importance of ICT standards for the Arabic language. It also provides information on major regional and international players involved in the Arabization process. Major problems, international standards and Internet protocols are also examined.

Some of the problems associated with Arabic and the Internet, including the existence of multiple character sets, can potentially be resolved by the adoption of Unicode. In addition, Unicode has the potential to alleviate other problems related to specialized display routines for Arabic, including the need for contextual analysis to select the appropriate character shapes and the incorporation of bi-directional displays.

The two most important Internet applications, namely, the Web and e-mail must operate flawlessly in Arabic. While interim solutions are available in this regard, a simpler and more satisfying solution must be devised. It is anticipated that this issue will be resolved when Web servers adhere to HTTP standards and include character set information in the header for page transmission and when Web browsers use that information to set up pages. The ability to search and index Arabic content on the Internet is crucial.

The Arab world can play a vital role in the ICT sector, which is increasing in importance, particularly with regard to products catering to the Arabic language. It must also be noted that the issue of standards for Arabic in ICT is more pressing at the regional, rather than the national level and therefore regional efforts must be accelerated accordingly. AIDMO-CSM must be supported and its endeavours enhanced.

In this context, the following recommendations are proposed:

(a) NSBs in the Arab world must form ICT technical committees, which will bear responsibility for issues related to the use of Arabic language in ICT systems. These committees will have the means to cooperate with TC-8 of AIDMO-CSM;

(b) NSBs in the Arab world must adopt a regional and modular project to coordinate their activities. The project must endeavour to unify procedures at all levels and with regard to all activities of NSBs. The project will enable NSBs to connect to the Internet and to network, by linking their sites, establishing databases of standards at each NSB and making catalogues available on the Internet, thereby synchronizing efforts and avoiding repetition of work. This project will also link NSBs to institutions in all countries, namely, chambers of commerce and industries, Ministries and R&D institutes;

(c) AIDMO and ESCWA must cooperate with other international standardization organizations to elaborate a standardization policy for the ESCWA region and the Arab world, and consequently, to issue recommendations for adopting standards particularly in the areas of IT, trade facilitation and quality management. Such a policy must aim to promote, adopt and harmonize standards, and must also strengthen the institutional capacity of NSBs, particularly in terms of activities in the areas of implementation, accreditation and certification;

(d) ESCWA member countries and the Arab region as a whole, must become increasingly aware of the need for standardization, and therefore, curricula of local education systems must include courses on standardization and quality management at all levels;

(e) The participation of the private sector in the activities of standardization is vital and must be encouraged.

Legislative and regulatory issues for information knowledge management are examined in the context of the absence of ICT Arabic terminology and standards for the utilization of Arabic language on the Internet and in relation to security issues.

Finally, this study proposes an action plan, which aims to contribute to the harmonization of ICT standards related to Arabic language use in information society applications. ESCWA member and Arab countries must therefore consider the following:

(a) *Formation of a Consulting Council*

This must be composed of an Arab international work team, including representatives of Arab countries and ITU, WTO, UNCTAD, ESCWA, AIDMO-CSM's TC-8 and AKMS. The Council will offer advice to governmental and private organizations in all Arab countries with regard to establishing the following:

(i) A unified Arab information network that will enable Arab communities to reap the benefits of technology;

(ii) Arab-consulting organizations that will cooperate with similar international organizations;

(iii) An Arab organization for integrated solutions that will service the Arab knowledge-based society on the basis of international standards.

(b) *Formation of a Technical Committee*

These must be components of the above-mentioned Consulting Council. One committee must endeavour to formalize a strategy for reducing ICT vulnerability that addresses legislative and regulatory issues for IKM and a security policy for the use of Arabic language on the Internet.

(c) *Formation of Technical Committee*

This technical committee must be a component of the above-mentioned Consulting Council and must endeavour to unify ICT terminologies in all ICT-related Arabic language standards.

(d) *Formation of an Internet Best Practices Committee*

This must also be a component of the above-mentioned Consulting Council for the Arabic language and must seek to develop a standard for web page engineering, which will address the following issues: copyright, proprietary data declarations, indexing and content classification of pages, multinational sensitivities, browser tolerance, bandwidth efficiencies, server operations and privacy.

(e) *Strengthening the role of TC-8 of AIDMO-CSM*

This will realize the objectives of the League of Arab States.

(f) *Formation of ICT technical committees by NSBs in the Arab world*

These committees must be responsible for issues related to the use of Arabic language in ICT systems and will cooperate with TC-8 of AIDMO-CSM. Such committees must consider the following measures as components of a strategy to reduce ICT vulnerability: establishing centres for information assurance; promoting increased R&D; improving knowledge and education; encouraging more risk and vulnerability analysis; employing user-targeted measures; and implementing legislation to combat crime.

(g) *Creation of policies to support R&D efforts*

Such efforts must be exerted in areas that include natural language processing.

(h) *Establishment of a security standard*

This must be related to security requirements and must be capable of facilitating the evaluation and measurement of security systems.

(i) *Adoption of Arabic language standards for the various aspects of storing and accessing digital information*

These aspects must include interoperability, data format, resource identification, resource description, archiving of data and records management.

(j) *Systematization*

All bodies, organizations, committees and technical sub-committees must be systemized into some kind of an operational hierarchy to avoid repetition and promote efficiency in the task of establishing all relevant ICT standards in the Arab world.

(k) *Promotion of OSS*

This can be achieved by doing the following:

(i) Encouraging the integration of courses related to open source concepts and philosophy in IT-related curricula at universities in the region;

(ii) Promoting investment in Linux and OSS in the public and private sectors at the regional level;

(iii) Raising awareness of Linux in the public and private sectors at the regional level.

Annex I

UNICODE

Unicode, or the Universal Multiple-Octet Coded Character Set (UCS), is a global character standard that is used for representing and encoding text to support the interchange, processing and presentation of written texts pertaining to multiple languages.[1] All Unicode codes are described by a descriptor prefixed by "U+" and followed by a hexadecimal number, which corresponds to the character position in Unicode. For instance, U+0661 is the Unicode character 'ARABIC-INDIC DIGIT ONE' and U+0000 to U+00FF covers the range of Latin-1 character set. The Unicode standard also includes punctuation marks, diacritics, mathematical symbols, technical symbols, arrows and dingbats. Unicode is fully compatible with the international standard ISO/IEC 10646-1, and contains the same characters and encoding as ISO/IEC 10646. Approximately 65,000 characters can be encoded in Unicode, which is sufficient for representing most character sets, including Arabic, Cyrillic, Latin, Thai and Tibetan. Furthermore, the Unicode standard and ISO 10646 can be extended to the so-called Unicode Transfer Format (UTF-16), which permits the encoding of a million more characters, without use of escape codes.

Before Unicode was invented, there were hundreds of different conflicting encoding systems to cover all languages. Unicode provides a unique number for every character regardless of the platform, programme and language used. It has been adopted by industry leaders such as Apple, HP, IBM, Microsoft, Oracle, Sun and Unisys, and is utilized by modern standards such as XML, Java, ECMAScript (JavaScript), LDAP, CORBA 3.0 and WML. Unicode is supported in many operating systems, all modern browsers, and many other products.[2] Incorporating Unicode into client-server or multi-tiered applications and web sites is a means of ensuring significant cost savings. Unicode enables a single software product or a single web site to be targeted across multiple platforms, languages and countries, and it allows data to be transported across many different systems without corruption.

A. THE UNICODE CONSORTIUM

The Unicode Consortium is a non-profit organization, which was founded to develop, extend and promote the use of Unicode.[3] Unicode provides a new foundation for the process of internationalization. Older code pages are more difficult to use, and have inconsistent definitions for characters. Internationalizing a code whilst using the same code base is complex, in that this process entails a need to support different character sets, with different architectures, for different markets. However, Unicode can produce code that handles the requirements of all world markets simultaneously. Given that Unicode has a single definition for each character, users do not get data error problems that plague mixed code set programmes. Moreover, given that Unicode handles the characters for all world markets in a uniform way, it avoids the complexities of different character code architectures. All modern operating systems, from PCs to mainframes, either support Unicode or are actively developing support for it. The same is true of databases. Unicode covers all languages that can be written in the following scripts: Latin; Greek; Cyrillic; Armenian; Hebrew; Arabic; Syriac; Thaana; Devanagari; Bengali; Gurmukhi; Oriya; Tamil; Telegu; Kannada; Malayalam; Sinhala; Thai; Lao; Tibetan; Myanmar; Georgian; Hangul; Ethiopic; Cherokee; Canadian-Aboriginal Syllabics; Ogham; Runic; Khmer; Mongolian; Han (Japanese, Chinese, Korean ideographs); Hiragana; Katakana; Bopomofo and Yi.

Unicode encodes characters on a per script basis. This means that, for example, only one set of Latin characters is defined, despite the fact that the Latin script is used for the alphabets of many languages. The same principle applies to other scripts, including, Arabic, which are used for writing a number of languages. However, Unicode does not encode scripts per se.[4]

---

[1] Unicode. Available at: http://www.unicode.org.

[2] Unicode, "Unicode enabled products". Available at: http://www.unicode.org/unicode/onlinedat/products.html.

[3] Unicode, "The Unicode Consortium". Available at: http://www.unicode.org/unicode/consortium/consort.html.

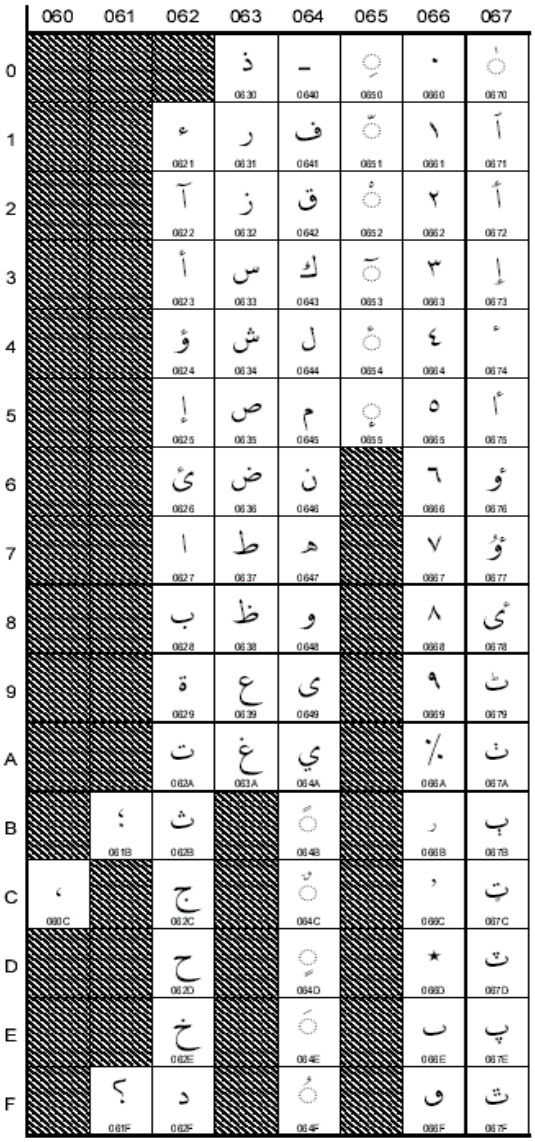
[4] A listing of script names is available at the Unicode web site (see footnote 105).

B. ARABIC CHARACTER CODE TABLES

The Arabic character code range in Unicode is: 0600–06FF (see annex tables 1 and 2).[5] The shapes of the reference glyphs used in code charts are not prescriptive and there is considerable variation in actual fonts, which were given to the Unicode Consortium by font designers, who own the rights to those fonts.[6]

Information on characters that have been added to Unicode since the publication of Version 3.2 is available at the Unicode web site;[7] information on characters being considered for addition can also be accessed through the same web site.[8]

ANNEX TABLE 1. THE ARABIC CHARACTER CODE RANGE 0600-067F IN THE UNICODE STANDARD



Source: Unicode, "Arabic range 0600-06FF". Available at: http://www.unicode.org/charts/PDF/U0600.pdf.

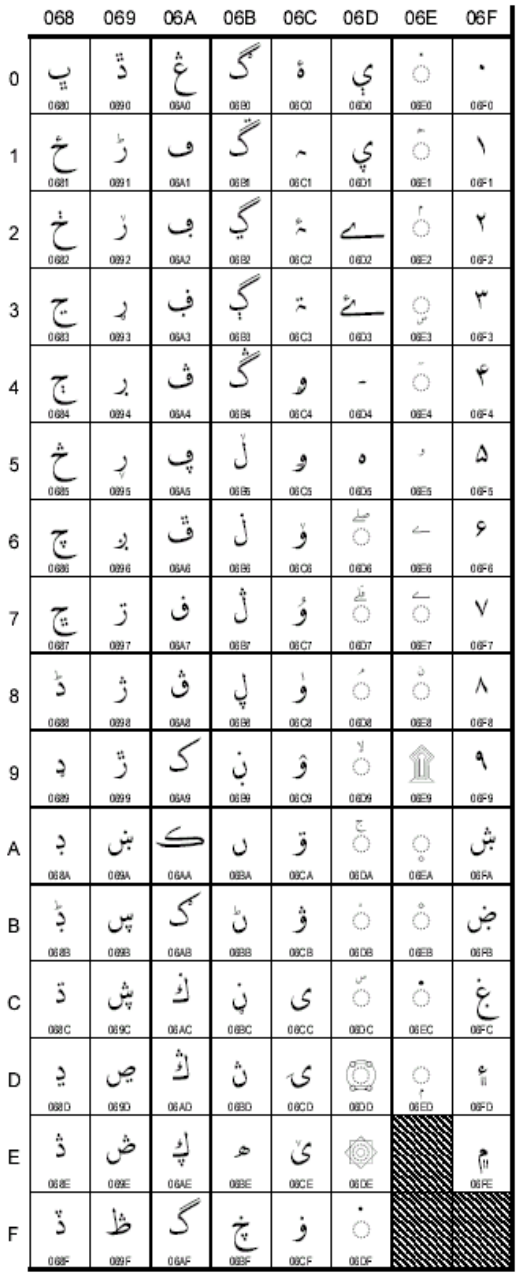
[5] Unicode, "Updates and errata". Available at: http://www.unicode.org/unicode/uni2errata/UnicodeErrata.html.

[6] Unicode, "Font acknowledgements". Available at: http://www.unicode.org/unicode/uni2book/u2fonts.html.

[7] Unicode, "Proposed new characters: pipeline table". Available at: http://www.unicode.org/unicode/alloc/Pipeline.html.

[8] Unicode, "Proposed new scripts". Available at: http://www.unicode.org/pending/pending.html.

ANNEX TABLE 2. THE ARABIC CHARACTER CODE RANGE 0680-06FF
IN THE UNICODE STANDARD



Source: Unicode, "Arabic range 0600-06FF". Available at: http://www.unicode.org/charts/PDF/U0600.pdf.

Annex II

ARABIC CODE TABLES

This annex contains a selection of Arabic code pages.[9]

ANNEX TABLE 3. MICROSOFT WINDOWS CODEPAGE 1256 (ARABIC)

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL 0000	STX 0001	SOT 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
10	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
20	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
30	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
40	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
50	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
60	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
70	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	| 007C	} 007D	~ 007E	DEL 007F
80	€ 20AC	پ 067E	‚ 201A	ƒ 0192	„ 201E	… 2026	† 2020	‡ 2021	ˆ 02C6	‰ 2030	ٹ 0679	‹ 2039	Œ 0152	ڤ 0686	ژ 0698	ڈ 0688
90	گ 06AF	‘ 2018	’ 2019	“ 201C	” 201D	• 2022	– 2013	— 2014	ک 06A9	™ 2122	ڑ 0691	› 203A	œ 0153	ZWNJ 200C	ZWJ 200D	ں 06BA
A0	NBSP 00A0	، 060C	¢ 00A2	£ 00A3	¤ 00A4	¥ 00A5	¦ 00A6	§ 00A7	¨ 00A8	© 00A9	ھ 06BE	« 00AB	¬ 00AC	- 00AD	® 00AE	¯ 00AF
B0	° 00B0	± 00B1	² 00B2	³ 00B3	´ 00B4	µ 00B5	¶ 00B6	· 00B7	¸ 00B8	¹ 00B9	؛ 061B	» 00BB	¼ 00BC	½ 00BD	¾ 00BE	؟ 061F
C0	ہ 06C1	ء 0621	آ 0622	أ 0623	ؤ 0624	إ 0625	ئ 0626	ا 0627	ب 0628	ة 0629	ت 062A	ث 062B	ج 062C	ح 062D	خ 062E	د 062F
D0	ذ 0630	ر 0631	ز 0632	س 0633	ش 0634	ص 0635	ض 0636	× 00D7	ط 0637	ظ 0638	ع 0639	غ 063A	ـ 0640	ف 0641	ق 0642	ك 0643
E0	à 00E0	ل 0644	â 00E2	م 0645	ن 0646	ه 0647	و 0648	ç 00E7	è 00E8	é 00E9	ê 00EA	ë 00EB	ى 0649	ي 064A	î 00EE	ï 00EF
F0	ً 064B	ٌ 064C	ٍ 064D	َ 064E	ô 00F4	ُ 064F	ِ 0650	÷ 00F7	ّ 0651	ù 00F9	ْ 0652	û 00FB	ü 00FC	LTR 200E	RTL 200F	ے 06D2

Source: Microsoft, "Microsoft Windows codepage 1256 (Arabic)". Available at: http://www.microsoft.com/globaldev/reference/sbcs/1256.htm.

[9] The information in the present annex, in addition to similar material, can also be located on the web site entitled "Character sets and code pages at the push of a button". Available at: http://www.i18nguy.com/unicode/codepages.html.

ANNEX TABLE 4. IBM PERSONAL COMPUTER ARABIC CODEPAGE 00864

HEX DIGITS 1ST → 2ND ↓	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0		► SM590000	(SP) SP010000	0 ND100000	@ SM050000	P LP020000	` SD130000	p LP010000	° SM190000	β GB010000	(RSP) SP300000	٠ ND100001	¢ SC040000	ذ AD470000	ـ SM860000	ّ AX100004
-1	☺ SS000000	◄ SM630000	! SP020000	1 ND010000	A LA020000	Q LQ020000	a LA010000	q LQ010000	· SD630000	∞ SA450000	(SHY) SP320000	١ ND010001	ء AX300000	ر AR010000	ف AF010003	ً AX100000
-2	♪ SM930000	↕ SM760000	" SP040000	2 ND020000	B LB020000	R LR020000	b LB010000	r LR010000	• SA790000	ϕ GF010001	آ AA210002	٢ ND020001	آ AA210000	ز AZ010000	ق AQ010003	ن AN010000
-3	♫ SM910000	!! SP330000	# SM010000	3 ND030000	C LC020000	S LS020000	c LC010000	s LS010000	√ SA800000	± SA020000	£ SC020000	٣ ND030001	أ AA310000	س AS010003	ك AK010003	ه AH010000
-4	☼ SM690000	¶ SM250000	$ SC030000	4 ND040000	D LD020000	T LT020000	d LD010000	t LT010000	▒ SF150000	½ NF010000	¤ SC010000	٤ ND040001	ؤ AW310000	ش AS230003	ل AL010003	ة AH010004
-5	═ SF430000	§ SM240000	٪ SM020007	5 ND050000	E LE020000	U LU020000	e LE010000	u LU010000	⌠ SF100000	¼ NF040000	ٱ AA310002	٥ ND050001	ع AC470002	ص AS450003	م AM010003	ى AA020002
-6	╟ SF240000	▬ SM700000	& SM030000	6 ND060000	F LF020000	V LV020000	f LF010000	v LV010000	⌡ SF110000	≈ SA700000		٦ ND060001	ئ AY310000	ض AD450003	ن AN010003	ي AY010002
-7	╬ SF440000	↨ SM770000	' SP050000	7 ND070000	G LG020000	W LW020000	g LG010000	w LW010000	┼ SF050000	« SP170000		٧ ND070001	ا AA010000	ط AT450000	ه AH010003	غ AG310004
-8	╣ SF230000	↑ SM320000	(SP060000	8 ND080000	H LH020000	X LX020000	h LH010000	x LX010000	╪ SF090000	» SP180000	ا AA010002	٨ ND080001	ب AB010003	ظ AZ450000	و AW010000	ق AQ010000
-9	╗ SF410000	↓ SM330000	) SP070000	9 ND090000	I LI020000	Y LY020000	i LI010000	y LY010000	┬ SF060000	لأ AL320000	ب AB010000	٩ ND090001	ة AT020000	ع AC470003	ى AA020000	لآ AL220000
-A	╝ SF420000	→ SM310000	* SM040007	: SP130000	J LJ020000	Z LZ020000	j LJ010000	z LZ010000	╞ SF080000	لأ AL320003	ت AT010000	ف AF010000	ت AT010003	غ AG310003	ي AY010003	لآ AL220003
-B	╜ SF400000	← SM300000	+ SA010000	; SP140000	K LK020000	[SM060000	k LK010000	{ SM110000	┴ SF070000		ث AT470000	؛ SP140007	ث AT470003	¦ SM650000	ض AD450000	ل AL010000
-C	╖ SF250000	∟ SA420000	, SP080000	< SA030000	L LL020000	\ SM070000	l LL010000	| SM130000	┐ SF030000		، SP080007	س AS010000	ج AG230003	¬ SM660000	ع AC470004	ك AK010000
-D	╓ SF390000	↔ SM780000	- SP100000	= SA040000	M LM020000	] SM080000	m LM010000	} SM140000	└ SF010000	لا AL020000	ج AG230000	ش AS230000	ح AH450003	÷ SA060000	غ AG310002	ي AY010000
-E	╙ SF380000	▲ SM600000	. SP110000	> SA050000	N LN020000	^ SD150000	n LN010000	~ SD190000	┘ SF020000	لا AL020003	ح AH450000	ص AS450000	خ AH470003	× SA070000	غ AG310000	■ SM470000
-F	╘ SF260000	▼ SV040000	/ SP120000	? SP150000	O LO020000	_ SP090000	o LO010000	⌂ SM790000	┌ SF040000	، SM870000	خ AH470000	؟ SP150007	د AD010000	ع AC470000	م AM010000	

Code Page 00864

Source: IBM, "F.36 code page 00864 (IBM Personal Computer) Arabic". Available at: http://publib.boulder.ibm.com/cgi-bin/bookmgr/BOOKS/QB3AQ501/F.36?SHELF=&DT=19971201194621.

ANNEX TABLE 5. MICROSOFT WINDOWS OEM CODEPAGE 720 (ARABIC)

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL 0000	STX 0001	SOT 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
10	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
20	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
30	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
40	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
50	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
60	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
70	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	| 007C	} 007D	~ 007E	DEL 007F
80			é 00E9	â 00E2		à 00E0		ç 00E7	ê 00EA	ë 00EB	è 00E8	ï 00EF	î 00EE			
90		ً 0651	ٌ 0652	ô 00F4	¤ 00A4	ـ 0640	û 00FB	ù 00F9	ء 0621	آ 0622	أ 0623	ؤ 0624	£ 00A3	إ 0625	ئ 0626	ا 0627
A0	ب 0628	ة 0629	ت 062A	ث 062B	ج 062C	ح 062D	خ 062E	د 062F	ذ 0630	ر 0631	ز 0632	س 0633	ش 0634	ص 0635	« 00AB	» 00BB
B0	▒ 2591	▓ 2592	▓ 2593	┃ 2502	┤ 2524	╡ 2561	╢ 2562	╖ 2556	╕ 2555	╣ 2563	║ 2551	╗ 2557	╝ 255D	╜ 255C	╛ 255B	┐ 2510
C0	└ 2514	┴ 2534	┬ 252C	├ 251C	─ 2500	┼ 253C	╞ 255E	╟ 255F	╚ 255A	╔ 2554	╩ 2569	╦ 2566	╠ 2560	═ 2550	╬ 256C	╧ 2567
D0	╨ 2568	╤ 2564	╥ 2565	╙ 2559	╘ 2558	╒ 2552	╓ 2553	╫ 256B	╪ 256A	┘ 2518	┌ 250C	▉ 2588	▄ 2584	▌ 258C	▐ 2590	▀ 2580
E0	ض 0636	ط 0637	ظ 0638	ع 0639	غ 063A	ف 0641	µ 00B5	ق 0642	ك 0643	ل 0644	م 0645	ن 0646	ه 0647	و 0648	ى 0649	ي 064A
F0	≡ 2261	◌ً 064B	◌ٌ 064C	◌ٍ 064D	◌َ 064E	◌ُ 064F	◌ّ 0650	≈ 2248	° 00B0	∙ 2219	· 00B7	√ 221A	ⁿ 207F	² 00B2	■ 25A0	NBSP 00A0

Source: Microsoft, "Microsoft Windows OEM codepage: 720 (Arabic)". Available at: http://www.microsoft.com/globaldev/reference/oem/720.htm.

ANNEX TABLE 6. ISO CHARACTER SET 8859-6 (ARABIC)

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL 0000	STX 0001	SOT 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
10	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
20	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
30	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
40	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
50	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
60	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
70	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	| 007C	} 007D	~ 007E	DEL 007F
80																
90																
A0	NBSP 00A0				¤ 00A4								، 060C	00AD		
B0												؛ 061B				؟ 061F
C0		ء 0621	آ 0622	أ 0623	ؤ 0624	إ 0625	ئ 0626	ا 0627	ب 0628	ة 0629	ت 062A	ث 062B	ج 062C	ح 062D	خ 062E	د 062F
D0	ذ 0630	ر 0631	ز 0632	س 0633	ش 0634	ص 0635	ض 0636	ط 0637	ظ 0638	ع 0639	غ 063A					
E0	ـ 0640	ف 0641	ق 0642	ك 0643	ل 0644	م 0645	ن 0646	ه 0647	و 0648	ى 0649	ي 064A	ً 064B	ٌ 064C	ٍ 064D	َ 064E	ُ 064F
F0	ِ 0650	ّ 0651	ْ 0652													

Source: ISO, "ISO character set 8859-6 (Arabic)". Available at: http://www.microsoft.com/globaldev/reference/iso/28596.htm.

ANNEX TABLE 7. IBM WINDOWS ARABIC CODEPAGE 01256

HEX DIGITS 1ST → 2ND ↓	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0			(SP) SP010000	0 ND100000	@ SM050000	P LP020000	` SD130000	p LP010000		گ AG010009	(RSP) SP300000	° SM190000		ذ AD470009	à LA130000	ً AA070009
-1			! SP020000	1 ND010000	A LA020000	Q LQ020000	a LA010000	q LQ010000	پ AP010009	‘ SP190000	، SP080007	± SA020000	ء AX300009	ر AR010009	ل AL010009	ٌ AU070009
-2			" SP040000	2 ND020000	B LB020000	R LR020000	b LB010000	r LR010000	‚ SP260000	’ SP200000	¢ SC040000	² ND021000	آ AA210009	ز AZ010009	â LA150000	ٍ AI070009
-3			# SM010000	3 ND030000	C LC020000	S LS020000	c LC010000	s LS010000	ƒ SC070000	“ SP210000	£ SC020000	³ ND031000	أ AA310009	س AS010009	م AM010009	ّ AA050009
-4			$ SC030000	4 ND040000	D LD020000	T LT020000	d LD010000	t LT010000	„ SP230000	” SP220000	¤ SC010000	´ SD110000	ؤ AW310009	ش AS230009	ن AN010009	ô LO150000
-5			% SM020000	5 ND050000	E LE020000	U LU020000	e LE010000	u LU010000	… SV520000	• SM570000	¥ SC050000	µ SM170000	إ AA310409	ص AS450009	ه AH010009	ْ AU050009
-6			& SM030000	6 ND060000	F LF020000	V LV020000	f LF010000	v LV010000	† SM340000	– SS680000	¦ SM650000	¶ SM250000	ي AY310009	ض AD450009	و AW010009	ٓ AI050009
-7			' SP050000	7 ND070000	G LG020000	W LW020000	g LG010000	w LW010000	‡ SM350000	— SM900000	§ SM240000	· SD630000	ا AA010009	× SA070000	ç LC410000	÷ SA060000
-8			(SP060000	8 ND080000	H LH020000	X LX020000	h LH010000	x LX010000	^ SD150100		¨ SD170000	ٍ SD410000	ب AB010009	ط AT450009	è LE130000	ّ AX100009
-9			) SP070000	9 ND090000	I LI020000	Y LY020000	i LI010000	y LY010000	‰ SM560000	™ SM540000	© SM520000	¹ ND011000	ة AT020009	ظ AZ450009	é LE110000	ù LU130000
-A			* SM040000	: SP130000	J LJ020000	Z LZ020000	j LJ010000	z LZ010000				؛ SP140007	ت AT010009	ع AC470009	ê LE150000	ٔ AE050009
-B			+ SA010000	; SP140000	K LK020000	[SM060000	k LK010000	{ SM110000	‹ SP270000	› SP280000	« SP170000	» SP180000	ث AT470009	غ AG310009	ë LE170000	û LU150000
-C			, SP080000	< SA030000	L LL020000	\ SM070000	l LL010000	| SM130000	Œ LO520000	œ LO510000	¬ SM660000	¼ NF040000	ج AG230009	ـ SM860000	ى AA020009	ü LU170000
-D			- SP100000	= SA040000	M LM020000	] SM080000	m LM010000	} SM140000	ج AC210009	(ZW N-J) SP5300Z0	(SHY) SP320000	½ NF010000	ح AH450009	ف AF010009	ي AY010009	(LR) SP5500Z0
-E			. SP110000	> SA050000	N LN020000	^ SD150000	n LN010000	~ SD190000	ژ AZ210009	(ZW Join) SP5400Z0	® SM530000	¾ NF050000	خ AH470009	ق AQ010009	î LI150000	(RLL) SP5600Z0
-F			/ SP120000	? SP150000	O LO020000	_ SP090000	o LO010000				¯ SD310000	؟ SP150007	د AD010009	ك AK010009	ï LI170000	

Code Page 01256

Source: IBM, “F.93 code page 01256 (Windows) Arabic”. Available at: http://publib.boulder.ibm.com/cgibin/bookmgr/BOOKS/QB3AQ501/F.93?SHELF=&DT=19971201194621.

Annex III

ARABIC AND INTERNATIONAL ORGANIZATIONS

Several international organizations are involved in ICT standardization, namely, the International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), International Telecommunication Union (ITU) and the Internet Engineering Task Force (IETF). ISO and IEC have a joint technical committee on IT, which is known as JTC1. There are 18 major subcommittees within JTC1 and two independent working groups. A brief review of JTC1 and its subcommittees is provided below:

(a) Scope: standardization in the field of IT;

(b) Total number of ISO standards and detailed information sheets (DISs) related to JTC1 and its subcommittees: 1322;

(c) Number of ISO standards and DIS under the direct responsibility of the Secretariat: IC549;

(d) Participating countries: 27; observer countries: 35;

(e) Liaison: ISO/IC 37, 39, 46, 68, 145, 154, 171, 184, 204 and IEC/TC 45, IOU;

(f) The subcommittees of JTC1 include: SC1 (Vocabulary); SC2 (Coded character sets); SC6 (Telecommunications and information exchange between systems); SC7 (Software engineering); SC11 (Flexible magnetic media for digital data interchange); SC17 (Identification cards and related devices); SC22 (Programming languages, their environments and system software interfaces); SC23 (Optical disk cartridges for information interchange); SC24 (Computer graphics and image processing); SC25 (Interconnection of information technology equipment); SC26 (Microprocessor systems); SC27 (IT Security techniques); SC28 (Office equipment); SC29 (Coding of audio, picture, multimedia and hypermedia information); SC31 (Automatic identification and data capture techniques); SC32 (Data management and interchange); SC34 (Document description and processing languages); and SC35 (User interfaces). The SCs whose scope of work pertains to ICT standards related to the Arabic language include: SC1, SC2, SC6, SC25, SC27, SC29 and SC32.

In addition to JTC1, ISO also has eight technical committees related to IT, including: TC10 (Technical drawings); TC37 (Terminology); TC46 (Information and documentation); TC145 (Graphical symbols); TC154 (Processes, data elements and documents in commerce, industry and administration); TC171 (Document imaging applications); TC173 (Technical systems and aids for disabled persons); and TC215 (Health informatics).

The European Telecommunications Standards Institute (ETSI) also plays a major role in developing a wide range of standards and other technical documentation, and in doing so, contributes to global standardization in telecommunications, broadcasting and IT. In this context, the signing of a Memorandum of Understanding between ETSI and the Arab Telecommunications Council of Ministers of the League of Arab States in 2002 has boosted the development of telecommunications in Europe and the Arab States. In recognition of the importance of standards in the development of telecommunications and global harmonization, several major Arab telecom players have begun to sign up as associate members of ETSI.

A. Arab Industrial Development and Mining Organization

The Arab Industrial Development and Mining Organization (AIDMO) creates advanced databases for Arab States in the fields of industry, mining and standardization. It is also developing Arab Industrial Information Centers (ARIFONET), which comprise and connect sub-networks and focal points in Arab countries and in Arab and international centres and databanks. ARIFONET promotes the exchange of industrial and technological information among Arab countries and the rest of the world. This exchange is made possible by connecting information centres in the fields of industry, mining, standards and measurement; electrical energy associations; technology transfer centres; industrial research centres; and statistics bodies, to the main information centre and to other regional and international information centres through the Internet.

ARIFONET provides information on AIDMO and its new publications and events; the General Secretariat of the League of Arab States; and, *inter alia*, the following sub-networks:[10]

(a) Arab Industrial Information Centers;
(b) Arab Sub Contracting (Bourse) Network;
(c) Arab Standardization and Metrology Associations Network;
(d) Arab Industrial Research Centers Network;
(e) Arab Electrical Energy Associations Network;
(f) Related Associations and Organizations Network;
(g) Chambers of Commerce and Industry Network.

AIDMO is endeavouring to achieve the following:

(a) Ensure Arab industrial coordination and integration;

(b) Contribute to the development, promotion and support of the Arab economy in the fields of industry, energy, mining and standardization with a view to strengthening productivity, quality and competitiveness;

(c) Plan for the support and elaboration of local, national and regional industrial projects and encourage investment in mining and industry in the Arab world;

(d) Establish Arab unified standards to facilitate trade among Arab countries;

(e) Promote technical, technological and industrial cooperation among Arab States, and with developed and developing countries.

The Center of Standardization and Metrology of AIDMO aims to unify standards among Arab States to promote inter-Arab trade. One of the objectives of the centre is to encourage Arab countries to cooperate in unifying standards, with the ultimate goal of adapting them to international standards. Furthermore, AIDMO includes a High Consultative Committee of Standardization, whose mission is to do the following:

(a) Study and adopt unified Arab standardization projects, and submit them to the Ministerial Council for final adoption. Over 1200 Arab standards have been issued;

(b) Propose policies and action plans pertaining to specifications, standards, laboratories, quality control and studies and submit them to the Executive Council for adoption;

(c) Set up, incorporate or abolish Arab standardization committees, adopt their work programmes and follow up on their implementation;

(d) Propose standardization training and on-going programmes;

(e) Propose policies geared towards the establishment and development of standardization entities in Arab States.

1. *Technical Committee 8*

The AIDMO-CSM Technical Committee 8 (TC-8) handles standards pertaining to the use of the Arabic language in ICT systems. TC-8 was formed in 1981 by the Arab Standards and Metrology Organization (ASMO) to deal with issues pertaining to Arabic characters in informatics. ASMO was later dissolved and replaced by CSM of AIDMO. The secretariat of TC-8 is affiliated to the Syrian Arab Organization for Standardization and Metrology (SASMO). TC-8 endeavours to achieve the following:[11]

[10] AIDMO. Available at: http://www.arifonet.org.ma/aidmo_us/main.htm.

[11] M. Mrayati, "Standards for Arabic in information technology", a paper presented at the Meeting on Standardization in the Arab Countries, Amman, 2-5 February 1999.

(a) Prepare standards for the use of Arabic in IT;

(b) Coordinate with other standardization organizations, namely, the European Computer Manufacturing Association, Arabic Task Force (ECMA-ATF);

(c) Participate at meetings of international technical committee and working groups on IT;

(d) Cooperate with R&D institutes and universities that are active in the area of IT.

TC-8 has held 12 official meetings since it was formed in 1981, and participated in several meetings held by ECMA-ATF, ISO and ALECSO. Standards developed by TC-8, some of which have been adopted as ISO standards are illustrated in annex table 8. Standards used in operating systems are detailed in annex table 9.

ANNEX TABLE 8. STANDARDS DEVELOPED BY TC-8

Arabic standard	Year	Description	ISO standard
ASMO-445	1982	Bilingual 5-bit code (telex)	
ASMO-449	1985	7-bit Arabic character-set SASO-429 (1986), GSMO-50	ISO/9036 (1987)
ASMO-584	1985	Trans-coding ASMO-449/ASMO-445	
ASMO-662	1985	8-bit Arabic character-set	
ASMO-663	1987	Arabic keyboard layout GSMO-596(1995)	
ASMO-708	1986	8-bit Arabic/Latin character-set. SASO-1139(1996), GSMO-653(1996)	ISO/8859-6 (1987)
ASMO-968	1988	Trans-coding ASMO-662/ASMO-445	
CSM-969	1992	Displayed and printed Arabic character-set	
CSM-1021	1992	Trans-coding AASMO-708/ASMO-445	
		16-bit Multilingual character-set	Unicode
		32-bit Universal Character Set (UCS)	ISO/10646

Source: M. Mrayati, "Standards for Arabic in information technology", a paper presented at the Meeting on Standardization in the Arab Countries, Amman, 2-5 February 1999.

ANNEX TABLE 9. STANDARDS USED IN OPERATING SYSTEMS

Operating System	Computer	Standard code
MS-DOS	IBM-PC and comp	Different non standard Codes
Windows-95	IBM-PC and comp	MS CP 1256
Windows-NT	IBM-PC and Minis	Unicode
Mac-OS	Macintosh	Modified ASMO-708
Arabized UNIX	Several	ASMO-708
VM	IBM Mainframe	XBASIC

Source: M. Mrayati, "Standards for Arabic in information technology", a paper presented at the Meeting on Standardization in the Arab Countries, Amman, 2-5 February 1999.

The present agenda of TC-8 includes the following standardization activities that are relevant to the use of Arabic in ICT systems:

(a) Participation in working groups, subcommittees and technical committees related to multilingual IT standards, including those that deal with the following issues:

(i) Use of FTP in Arabic;

(ii) Use of Arabic site-names on the Internet;

(iii) Use of Arabic in e-mail: SMTP (SMTP, RFC821); RFC822; MIME, parts 1, 2, 3 and 4 (RFC2045, 2046, 2047 and 2048);

(iv) Internet protocols and languages dealing with multilingual environment: transfer protocol HTTP 1.1; markup languages: HTML 2.0, 4.0, in coordination with RFC-2070; and XML.

(b) Participation in the finalization of the Arabic pages of the ISO/10646 character-set standard, and adoption of the results of this finalization at the AIDMO-CSM level;

(c) Adoption of an Arabic version of United Nations Electronic Data Interchange for Administration, Commerce and Transport UN/EDIFACT, ISO/9735, the standard used in e-commerce for electronic data interchange (EDI) and the new standard for electronic form interchange (EFI);

(d) Development of multilingual terminology, transliteration and localization standards;

(e) Testing, accreditation and certification procedures and standards.

B. The Arab Knowledge Management Society

The Arab Management Society (AMS) was established in 1989. The original objective of AMS was to improve management in Arab countries to increase productivity. AMS became the Arab Knowledge Management Society (AKMS) to reflect the importance of the ability to manage and utilize knowledge in the modern age. [12] AKMS aims to utilize modern management and technology to develop Arab social and economic capabilities. AKMS also seeks to encourage the development of the Arab knowledge society (see annex box 1).

Annex box 1. Goals of the Arab Knowledge Management Society

The AKMS aims to do the following:

(a) Focus on the social, cultural and organizational nature of knowledge management within the unique context of the Arab social and business culture;

(b) Contribute to Government policy debates on proper measures for the management and development of the technical infrastructure of the knowledge society;

(c) Comprehensively define and explain the meaning of knowledge management and disseminate this understanding to Arab businessmen so that they can more fully utilize knowledge management principles and identify knowledge management business opportunities;

(d) Study the social, cultural, economic, organizational and technological issues that have an impact on the effectiveness of knowledge management initiatives and existing systems in Arab countries;

(e) Study the social and cultural factors that encourage individuals to share knowledge within an organization;

(f) Increase the number of people with access to computers, the Internet and related technologies;

(g) Formulate pan-Arab business statements on knowledge society policies;

(h) Contribute an Arab business perspective to multilateral forums concerned with issues related to the global knowledge society;

(i) Act as a representative of the pan-Arab business community before League of Arab States regarding issues concerning the knowledge society;

(j) Develop an Arab Knowledge Management Certification system, which will contribute to the development of professional and educational standards for knowledge management professionals.

Source: AKMS. Available at: http://www.akms.org/goals.htm (in Arabic).

[12] Arab Knowledge Management Society (AKMS). Available at: http://www.akms.org.

C. Talal Abu-Ghazaleh International Society

The Talal Abu-Ghazaleh International Society (TAGI), is involved in a variety of activities, which unite economic, social and cultural development in the Arab world in the context of a global economy. [13] TAGI and AKMS, in close cooperation with the International Chamber of Commerce have been endeavouring to develop e-business initiatives in the Arab world.[14] These three organizations, in coordination with the Ministry of Commerce and Industry of Oman, held a conference on Meeting the Challenges of Electronic Business in Muscat in 2000, at which a joint Arab e-business initiative, which was incorporated in the Muscat Declaration, called for the formation of an Arab international task force to promote the creation of a knowledge-based Arab economic environment and to foster innovation to enhance creativity and competition in the knowledge-based global economy. The role of the task force is to propose and implement business, technical and legal programmes in Arab countries with the objective of creating an enabling e-business environment.

D. Internet Corporation for Assigned Names and Numbers

The Internet Corporation for Assigned Names and Numbers (ICANN) is a non-profit corporation that was formed in 1998 to assume responsibility for the allocation of IP address space, coordination of the protocol number assignment, management of the domain name system (DNS) and management of the root server system functions that were previously performed under a United States Government contract by the Internet Assigned Number Authority (IANA) and other entities.[15] ICANN was created by members of the Internet community in response to a June 1998 White Paper, which was issued by the United States Department of Commerce (DOC).[16] ICANN represents an endeavour to create a globally representative private sector policymaking body.

E. Arabic Internet Names Consortium

The Arabic Internet Names Consortium (AINC) is an international organization, whose members include country code top-level domain (ccTLD) managers and organizations. AINC was founded to Arabize domain names and to facilitate Internet internationalization for all Arabic-speaking people. AINC aims to develop and deploy an Arabic domain name (ADN) system and related applications. A cooperation agreement between AINC and the Multilingual Internet Names Consortium (MINC) signed in 2001, effectively initiated the task of Arabizing domain names on the Internet.

F. Arab League Educational, Cultural and Scientific Organization

The League of Arab States is a regional organization that was established to strengthen ties between the 22 Arab States, coordinate policies and promote common interests. The League includes several specialized organizations and councils, namely, the Council of Arab Ministers Responsible for the Environment (CAMRE) and the Arab League Educational, Cultural and Scientific Organization (ALECSO). ALECSO was established in 1970 and its primary responsibility is the promotion and coordination of educational, cultural and scientific activities at the regional level.[17] It is involved in the creation and implementation of new approaches and strategies related to educational, cultural and scientific development in line with Arab realities, needs and priorities (see annex box 2).

---

[13] Also known as the Talal Abu-Ghazeleh Organization. Available at: http://www.tagi.com.

[14] Talal Abu-Ghazaleh International Society (TAGI) and Arab Knowledge Management Society (AKMS), "Towards an Arab knowledge society", a joint report. April 2001.

[15] The Internet Corporation for Assigned Names and Numbers (ICANN). Available at: http://www.icann.org/.

[16] United States Department of Commerce, "Management of Internet names and addresses", June, 1998. Available at: http://www.icann.org/general/white-paper-05jun98.htm.

[17] Arab League Educational, Cultural and Scientific Organization (ALECSO). Available at: http://www.slis.uwm.edu/alecso/.

Annex box 2. Objectives of the Arab League Educational, Cultural and Scientific Organization

The objectives of Arab League Educational, Cultural and Scientific Organization (ALECSO) include the following:

"(a) To work for the development of human resources in the Arab countries by helping the Arab individual to develop intellectually, morally and culturally to his maximum potential, so that he may share in the building of the Arab society and in the development of his own civilization and that of mankind;

"(b) To share in the development of Arab economic and social sciences and their application to issues daily affecting the development of the standard of living in the Arab world;

"(c) To effect progress in the factors of scientific development and the application of modern technology, with the purpose of creating the appropriate atmosphere for the transfer of modern technology into the Arab countries;

"(d) To promote the improvement of environmental development in the Arab countries with the purpose of ensuring a proper scientific use of their resources consistent with their present and future requirements of the Arab countries;

"(e) To explore new areas of linking the Arab-Islamic thought with contemporary experience in the context of crucial issues facing the world, and to work for the establishment of peace in its broad concept as well as safeguarding human rights for the setting up of an equitable new international order consistent with Arab-Islamic values;

"(f) To develop Arab-Islamic culture at home and abroad within the context of global knowledge, and to contribute positively in the establishment of the essentials of an international culture which focuses on the great human values of the world societies;

"(g) To seek the most effective ways and forms of participation in international endeavours for the development of mass media and communication as well as data processing and documentation to facilitate their circulation and render them more democratic in the service of the objectives of the Organization".

Source: Arab League Educational, Cultural and Scientific Organization (ALECSO), "Objectives and purposes/priorities". Available at: http://www.slis.uwm.edu/alecso/alecsopurps.htm.

G. Arab Club for Information

Founded in 1998, the Arab Club for Information (Arabcin) is a scientific cooperative non-governmental Arab institution, which is endeavouring to devise a joint Arab strategy to face the challenges of the information age.[18] Arabcin primarily promotes the generation, digitization and dissemination of Arabic content. It also strengthens cooperation and coordination among Arab institutions in the field of information, documents and archives (see annex box 3).

Annex box 3. Arabcin institutions

Arabcin institutions are briefly reviewed below:

(a) *Arab Research and Study Center*

This entity conducts studies and carries out research, and is responsible for organizing assemblies, seminars, interviews and related workshops. It has plans to form a scientific board of advisors including Arab researchers;

(b) *Arab Biographic Center*

This body collects and preserves biographies of Arabs employed in all fields in the form of a database.

[18] Arab Club for Information (Arabcin). Available at: http://www.arabcin.net/english/index.htm.

Annex box 3 (*continued*)

(c) *Arab Institute for Information*

This provides training courses on information, documentation, archives, IT, communications and libraries to improve the level of Arab expertise in these areas; promotes cooperation among experts and trainers; and develops methods to facilitate the exchange of information among club members.

(d) *Arab Institution for ICT*

This institution develops and maintains the ICT infrastructure of Arabcin;

(e) *Arab Institution for Press, Printing* (also known as the *Information Institution for Press, Publication and Distribution)*

This entity deals with typing; technical design; colour separation; montage and press; binding; publication; and marketing.

Source: Arabcin, "Institutions". Available at: http://www.arabcin.net/english/institutes.htm.

H. The Bahrain Directorate of Standards and Metrology

The Bahrain Directorate of Standards and Metrology (BSMD) was established in 1975 and is a Government Directorate of the Ministry of Commerce. This NSB is a member of the following standardization organizations:

(a) Gulf Standards and Metrology Organization (GSMO);
(b) ISO;
(c) National Standard Body of Egypt.

I. Egyptian Organization for Standardization and Quality Control

The Egyptian Organization for Standardization and Quality Control (EOS) was established in 1957 and is a NSB that is affiliated to the Ministry of Industry and Technology. EOS is responsible for the creation of standards and related activities; it also coordinates standardization activities at the national level and with Arab, foreign, and international organizations in the fields of standardization and metrology. EOS aims to do the following:

(a) Develop, adopt, publish, amend and revoke Egyptian national standards and codes of practice for raw materials; industrial products services, testing and measuring equipment; methods of testing and inspection; quality control; calibration; and metrology;

(b) Perform testing and inspection services;

(c) Grant conformity certification of products and systems;

(d) Grant quality marks of Egypt;

(e) Represent the country in, and/or coordinate with, related Arab, foreign and international standards bodies or organizations;

(f) Offer technical consultation and training services in the field of standardization.

EOS is a member of a number of standardization and related bodies, including the following:

(a) AIDMO-CSM, as a participating member;
(b) The African Regional Organization for Standardization (ARSO), as a participating member;
(c) ISO, as a full member;
(d) IEC, as a participating member;
(e) The International Organization of Legal Metrology (OIML), as a participating member;
(f) The International Federation for the Application of Standards (IFAS), as a participating member;
(g) Codex Alimentarius Commission (Codex), as a participating member.

EOS also cooperates, coordinates and exchanges standards and publications with the following:

(a) British Standards Institution (BSI);
(b) Association francaise de normalisation (AFNOR);
(c) American National Standards Institute (ANSI);
(d) Saudi Arabian Standards Authority (SASO).

J. IRAQI CENTRAL ORGANIZATION FOR STANDARDIZATION AND METROLOGY

The Iraqi Central Organization for Standardization and Metrology (ICOSM) is a government body that was established in 1963. ICOSM was transformed into an independent agency affiliated with the Ministry of Planning in 1979.

K. JORDANIAN INSTITUTION FOR STANDARDS AND METROLOGY

The Jordanian Institution for Standards and Metrology (JISM) is a government body that is administratively and financially independent and is affiliated with the Ministry of Industry and Trade. JISM is a NSB that is responsible for adopting a national system for standards and metrology; keeping pace with scientific developments in its field of responsibility; providing health and environmental protection for citizens, by ensuring conformity of commodities and other materials to approved standards; and supporting the national economy and economic development plans, by assuring quality and competitiveness of locally produced products. JISM aims to do the following:

(a) Prepare, approve, revise, amend, replace and monitor the implementation of standards for commodities and materials with the exception of drugs, veterinary medicines, serums and vaccines. The institution may accept and approve standards of other countries and of Arab, regional and international organizations, provided that such standards are in Arabic or English;

(b) Establish a national system of measurement and supervise its implementation; unify means and methods of measurement; develop, calibrate, adjust and control measuring instruments; and approve the national basic measurement standards necessary for calibrating measuring;

(c) Develop instruments for hall-marking or stamping;

(d) Support and promote studies and research at accredited testing laboratories, in fields relevant to standards, metrology and quality control; organize relevant training courses; and publish, distribute and sell publications related to approved standards and other related subjects;

(e) Grant quality marks and certificates of conformity; and accredit qualified testing or calibration laboratories at public and scientific institutions;

(f) Cooperate and coordinate with, or join the membership of, Arab, regional and international organizations active in the fields of standardization and metrology; and conclude mutual recognition of quality marks and certificates of conformity with these bodies.

JISM is a member of the following standardization bodies/organizations:

(a) AIDMO-CSM, as a participating member;
(b) ISO, as a correspondent member;

(c) OIML, as a correspondent member;
(d) Codex, as a contact point.

L. KUWAIT STANDARDS AND METROLOGY DEPARTMENT

The Kuwait Standards and Metrology Department (KSMD) was established in 1977 and is a department of the Ministry of Commerce and Industry. It is a NSB and is responsible for creating standards and related activities, and coordinating standardization activities at the national level and with Arab, foreign, and international organizations in the fields of standardization and metrology.

KSMD aims to do the following:

(a) Develop, publish, propagate and modify national standards covering raw materials, commodities and goods;

(b) Promote, follow-up and monitor the use of national Kuwaiti standards;

(c) Issue, disseminate and sell standards and standardization related publications;

(d) Keep primary and national legal metrology standards references and issue calibration certificates;

(e) Establish an information and documentation centre to maintain and disseminate information and statistics related to standardization;

(f) Provide the means to verify and ensure conformity to national or nationally approved standards for raw materials, commodities and goods;

KSMD is a member of the following standardization bodies:

(a) GSMO;
(b) AIDMO-CSM, as a participating member;
(c) ISO, as a correspondent member;
(d) OIML;
(e) Codex.

M. LEBANESE STANDARDS INSTITUTION (LIBNOR)

The Lebanese Standards Institution (LIBNOR) was established in 1962 and is independent in terms of finances and administration. LIBNOR aims to do the following:

(a) Develop, publish, and modify national standards in areas including, but not limited to: measurements, terms and symbols; specifications for quality of goods and associated inspection and testing procedures; codes of professional practice and technical rules in construction;

(b) Grant the right to use the symbol of quality mark, to indicate conformity of products to relevant national standards;

(c) Represent the country, in the field of standardization.

LIBNOR is a member of the following standardization bodies:

(a) AIDMO-CSM, as a participating member;
(b) ISO, as a correspondent member.

N. Directorate General For Specifications and Measurements of Oman

The Directorate General for Specifications and Measurements of Oman (DGSM) was established in 1976; it is a department of the Ministry of Commerce and Industry and is NSB. DGSM is responsible for creating standards and related activities and coordinating standardization activities at the national level and with Arab, foreign, and international organizations in the fields of standardization and metrology. DGSM aims to do the following:

(a) Prepare, publish, distribute and update standards for commodities; products; raw materials; methods of sampling and testing; measures; definitions and symbols;

(b) Promote awareness of standardization among all concerned authorities and the public;

(c) Disseminate standards and standardization-related publications at the national level;

(d) Organize legal and industrial metrology in Oman;

(e) Provide inspection and testing services to verify and ensure conformity of various materials and commodities to relevant standards;

(f) Conduct necessary testing for precious metals, to identify carat contents, and inspect gemstones;

(g) Issue export validity certificates for foodstuffs destined for export to other GCC countries;

(h) Adopt Gulf standards, namely, GSMO standards as the standards of Oman;

(i) Coordinate national standardization activities with those of regional and international standards bodies;

(j) Participate in conferences and meetings and represent Oman at both the regional and international levels;

(k) Sell Omani and GSMO standards.

DGSM is a member of the following standardization bodies:

(a) GSMO, as a participating member;
(b) AIDMO-CSM, as a participating member;
(c) ISO, as a correspondent member;
(d) OIML, as a correspondent member.

O. Palestinian Standards and Specifications Establishment

The Palestinian Standards and Specifications Establishment (PSSE) is independent in terms of finances and administration. It is responsible for creating standards and related activities and coordinating standardization activities at the national level and with Arab, foreign, and international organizations in the fields of standardization. PSSE aims to do the following:

(a) Develop, publish, revise, replace and control implementation of national standards for goods, and materials;

(b) Develop and grant quality mark, and issue conformity certification, to indicate conformity of products to relevant national standards.

P. Qatar Department of Standards, Measurements and Consumer Protection

Qatar Department of Standards, Measurements and Consumer Protection (QDSMCP) was established in 1972 and is a Department of the Ministry of Finance, Economy and Commerce. This NSB is responsible

for creating standards and related activities and coordinating of standardization activities at the national, regional and international levels. QDSMCP aims to do the following:

(a) Develop and publish Qatari national standards;

(b) Follow-up related standards work of the General Secretariat of GSMO;

(c) Implement national laws, rules, directives and regulations related to the national standards and measurements system;

(d) Verify the accuracy of information in commercial labels and its adherence to the required labelling standards;

(e) Issue the certificates of conformity for goods produced locally, imported or exported; and grant the quality mark for locally produced products;

(f) Represent the Ministry in all standardization activities conducted by international organizations in the fields of standardization and metrology;

(g) Conduct R&D related to local production in line with approved standards, in coordination with the relevant national bodies;

(h) Implement national laws, rules, directives and regulations related to the unification of national measurements, volumes and weights;

(i) Implement national laws, rules, directives and regulations related to: (a) inspecting, testing and stamping of precious metals; and controlling prices/profits and swindling in commercial transactions, and (b) verifying abidance with consumer protection laws and regulations.

QDSMCP is a member of the of the following standardization bodies:

(a) GSMO, as a participating member;
(b) AIDMO-CSM, as a participating member;
(c) ISO, as a correspondent member.

Q. Saudi Arabian Standards Organization

The Saudi Arabian Standards Organization (SASO) has judicial powers and an independent budget.[19]

SASO aims to do the following:

(a) Formulate and approve national standards for all commodities and products, and standards concerned with metrology, calibration, marking and identification of commodities and products, methods of sampling, inspection and testing;

(b) Publish Saudi Arabian standards;

(c) Promote awareness of standardization and coordinate all activities relating to standards and measurements in the country;

(d) Outline rules for granting certificates of conformity and quality mark and regulating their issuance and use;

(e) Participate in Arab, regional and international organizations.

[19] The Saudi Arabian Standards Organization (SASO), "Objectives". Available at: http://www.saso.org.sa/english/information/objectives/information.htm.

R. SYRIAN ARAB ORGANIZATION FOR STANDARDIZATION AND METROLOGY (SASMO)

The Syrian Arab Organization for Standardization and Metrology (SASMO) was established in 1969. In 1980, SASMO became administratively independent, whilst remaining affiliated with the Ministry of Industry. SASMO is responsible for creating standards and related activities and coordinating standardization activities at the national level and with Arab, foreign and international organizations in the fields of standardization and metrology. SASMO is the only body that issues standards in the Syrian Arab Republic, with two exceptions: (a) the Ministry of Health, which is responsible for human drug serums and vaccines; and (b) the Ministry of Agriculture, which is responsible for similar products for animals.

SASMO aims to do the following:

(a) Develop, publish, propagate and modify national standards: raw materials, commodities and goods including unified classifications, definitions, technical terms and symbols;

(b) Establish references for unified measurement and calibrate and maintain the required measuring instruments, equipment and other devices related to legal metrology;

(c) Verify the quality of raw materials, commodities and goods, and their conformity to adopted standards;

(d) Grant the rights to use the symbol of quality mark to indicate conformity of products to relevant national standards;

(e) Represent the country in Arab, foreign and international standards bodies;

(f) Hold symposiums, seminars, conferences and training courses on standardization, metrology, quality control and quality mark;

SASMO is a member of the following standardization bodies:

(a) AIDMO-CSM, as a participating member;
(b) ISO, as a full member;
(c) IEC, as a correspondent member;
(d) OIML, as a correspondent member;
(e) Food and Agriculture Organization (FAO), as a national contact point;
(f) American Society for Testing and Materials (ASTM), as a correspondent member;
(g) The Arab Union for Concrete and Building Materials (AUCBM), as a correspondent member;
(h) International Dairy Federation (IDF), as a correspondent member.

S. DIRECTORATE OF STANDARDIZATION AND METROLOGY OF THE UNITED ARAB EMIRATES

The Directorate of Standardization and Metrology of the United Arab Emirates (SSUAE) was established in 1976 and is affiliated to the Ministry of Finance and Industry, which is responsible for creating standards and related activities and coordinating standardization activities at the national level and with Arab, foreign, and international organizations, in the fields of standardization and metrology.

SSUAE aims to do the following:

(a) Develop, publish, and amend national standards for: units of measurements; classifications; terminology; definitions; codes; signs; dimensions, shapes; types, classifications, grades, patterns, quality, performance, durability and safety of materials, products, commodities, equipment, assemblies and constructions; methods of sampling, inspection, analyses, testing, measurements, calibration and controlling product quality; processing; canning; packing; transportation; distribution; storage; utilization; operation; and

for bases of engineering drawings, designs and tender conditions for projects, construction contracts and technical installations;

(b) Promote, follow-up and monitor the use of national United Arab Emirates standards in various national activities;

(c) Perform the following services in metrology:

(i) Retain custody of primary national reference standards for measurement;

(ii) Use these reference standards to inspect, compare, calibrate and organize the issuance of calibration certificates;

(iii) Issue, publicize, distribute and sell standards and other publications related to unification of measurements;

(iv) Generate awareness on the importance on unification of measurements, using all possible means, including education, training and raising the efficiency of personnel working at various levels in this field;

(d) Establish and operate laboratories and institutions for inspection, analysis, testing and research related to standardization and metrology; test materials and products; and calibrate measuring equipment;

(e) Grant conformity certifications and quality mark; and regulate the mode of issue and the right of use;

(f) Endeavour to harmonize national standards with those of other Arab countries and with other specialized organizations and agencies;

(g) Coordinate and cooperate with other bodies, organizations and institutions;

(h) Represent the country in conferences, organizations, bodies, committees and regional and international meetings.

SSUAE is a member of the following standardization bodies:

(a) GSMO, as a participating member;
(b) AIDMO-CSM, as a participating member;
(c) ISO, as a correspondent member.

T. GULF STANDARDS AND METROLOGY ORGANIZATION

GSMO was established in 1982 as a regional organization, for the GCC Countries. It is independent in terms of finances and administration and it is responsible for creating standards and related activities and coordinating standardization activities among its member countries.

GSMO aims to do the following:

(a) Develop, approve and publish Gulf States (GS) standards, for: goods; products; instruments of measurements and calibration; definitions; symbols; technical terms; conditions for execution; and procedures for sampling, inspection, testing and calibration;

(b) Prepare, print and publish GS in coordination with member countries;

(c) Follow-up on the implementation of GS through the NSBs of its member States;

(d) Develop a master plan for standardization activities in its member States in all fields, and distribute the burden of its implementation to all member States, based on their available resources;

(e) Organize legal and industrial calibration activities;

(f) Develop a system for granting quality marks and conformity certification;

(g) Conduct research and studies related to the development of the work of the organization;

(h) Generate awareness on standardization;

(i) Store information related to the work of the organization for the benefit of GCC countries;

(j) Develop and organize plans for training activities in various related fields of activities;

(k) Offer technical assistance to its member States;

(l) Obtain observing membership status in regional and international organizations that conduct relevant standardization activities.

Arabic and international organizations that deal with issues related to ICT Arabization standards are illustrated in annex table 10.

ANNEX TABLE 10. ARABIC AND INTERNATIONAL ORGANIZATIONS THAT DEAL WITH ASPECTS PERTAINING TO ICT ARABIZATION STANDARDS

		Organization	Comment
International	ISO	International Organization for Standarization	ISO has more than eight technical committees related to information technology (IT)
	IEC	International Electrotechnical Commission	ISO and IEC have a joint technical committee on IT, known as JTC1
	ITU	International Telecommunication Union	
	IETF	Internet Engineering Task Force	
	ICANN	Internet Corporation for Assigned Names and Numbers	Manages domain names
	ETSI	European Telecommunications Standards Institute	Provides standardization in telecommunications, broadcasting and IT
Arab	AIDMO	Arab Industrial Development and Mining Organization	Unifies standards among Arab States. Its TC-8 handles standards pertaining to the use of the Arabic language in ICT systems
	AKMS	Arab Knowledge Management Society	Utilizes modern management and technology to develop Arab social and economic capabilities
	TAGI	Talal Abu-Ghazaleh International Society	Fosters innovation and enhances creativity in the knowledge-based global economy
	AINC	Arabic Internet Names Consortium	Exerts efforts to internationalize the Internet and Arabize domain names
	ALECSO	Arab League Educational, Cultural and Scientific Organization	Promotes and coordinates educational, cultural and scientific activities at the regional level
	Arabcin	Arab Club for Information	Promotes generation, digitization and dissemination of Arabic content

ANNEX TABLE 10 (*continued*)

		Organization	Comment
Arab National Standardization Bodies (NSBs)	BSMD	Bahrain Directorate of Standards and Metrology	NSB in the Ministry of Commerce
	EOS	Egyptian Organization for Standardization and Quality Control	NSB in the Ministry of Industry and Technology
	GSMO	Gulf Standards and Metrology Organization	Regional organization for the GCC countries
	ICOSM	Iraqi Central Organization for Standardization and Metrology	NSB in the Ministry of Planning
	JISM	Jordanian Institution for Standards and Metrology	NSB in the Ministry of Industry and Trade
	KSMD	Kuwait Standards and Metrology Department	NSB in the Ministry of Commerce and Industry
	LIBNOR	Lebanese Standards Institution	NSB in the Ministry of Industry
	DGSM	Directorate General For Specifications and Measurements of Oman	NSB in the Ministry of Commerce and Industry
	PSSE	Palestinian Standards and Specifications Establishment	
	QDSMCP	Qatar Department of Standards, Measurements and Consumer Protection	NSB in the Ministry of Finance, Economy and Commerce
	SASO	Saudi Arabian Standards Organization	NSB in the Ministry of Industry
	SASMO	Syrian Arab Organization for Standardization and Metrology	NSB in the Ministry of Industry
	SSUAE	Directorate of Standardization and Metrology of the United Arab Emirates	NSB in the Ministry of Finance and Industry