
Nine problems concerning Arabic

**By Research Librarian / Charlotte Wien
Center for Contemporary Middle East Studies
Odense University, Denmark**

Introduction

Currently we have a number of systems that work with Arabic script. We have ASMO 449 and ASMO 708 and we have Unicode and ISO 10464. There is Windows in Arabic and Arabic language extensions for Mac, and even advanced database software and word-processing including spell checkers on the micro-computer level.

As a result of these developments, we can work with the written Arabic language in databases, word-processing etc. as if it was any language written in Latin letters. Of course, we also have integrated library systems that can write, save, print, retrieve and display records in Arabic script; this means that we can do the cataloging and indexing of Arabic documents in the Library databases as if they were just like all

the other documents - at least, that's what we believe!

For those who might not know ASMO, Unicode and ISO 10464 - I owe an explanation: ASMO is basically the same as ASCII, that means a unified character set especially for Arabic script. Unicode and ISO 10464 are character sets that contain all the different scripts of the world. Unified character sets are the basis of all data exchange. And in principle, if everyone was using the same character-set, all of us would be able to exchange records.

When I stated above that we have integrated library systems capable of handling Arabic script, I am referring mainly to five bibliographic database systems: The Canadian MINISIS (Beauvais, 1984), the Dutch DOBIS / LIBIS (Khurshid,

1987), the Israeli ALEPH(Panzer, 1992), the American system behind RLIN (Aliprand, 1992a) and according to IME in London, TINLIB is also capable of handling Arabic script¹. So really - what is the big problem? Unfortunately there are several. First of all, not all libraries are using these systems, secondly we cannot just add records in Arabic script to the existing on - line facilities and expect them to behave like records in, written in the Latin, Greek or Cyrillic alphabet. My task is to outline what makes the Arabic language work different in bibliographic databases.

I shall present basic examples in Arabic script and in order to make my paper comprehensive, I shall use the Library of Congress romanization standard along with the Arabic script.

In my introduction I have implicitly raised the question of whether bibliographic material in Arabic script can be treated just like records in any other language. The problems involved in using vernacular Arabic script can be divided into two categories:

1) Technical problems

2) "Linguistical" problems

It is a feature of the first group of problems that they are directly linked with the technical problems, within the computer environment (problem 1-4²). For the second group of problems these are indirectly linked with the computer environment. The second group of problems are problems which arise when the technical problems have been solved, and when records in vernacular Arabic script are added to bibliographic databases (problem 5-9³). In order to clarify the terminology and nature of the problems, I find it necessary to firstly, give a short outline of the differences between the written Arabic language and Indo-European languages like for example English. Secondly, I shall give an outline of the previous research in this area. Thirdly, I shall analyze which of the problems concerning the Arabic script in bibliographic databases still needs to be solved, and finally, I shall conclude whether the research in this area has brought the ability to retrieve records in

¹ This has not been proved yet.

² - Problem no. 1: The alphabet, Problem no. 2: The writing direction, problem no. 3: The graphical expression

³ - Problem no. 5: The lack of vowels, problem no. 6: Roots and patterns, problem no. 7: Weak radicals, problem no. 8: Case endings and pronouns, problem no. 9: Orthography.

Arabic script up to the same level as records in Latin script.

The written Arabic Language

Arabic is considered one of the "large" languages. It is the official language of 18 countries, having a total population of about 300 mio. people. Although Arabic is spoken in so many countries, there are huge differences between the written and the spoken language through the Arab world. In principle, the written language is the same for all countries⁴. This section will discuss problems related to the written language⁵.

The first noticable difference between written English and written Arabic is the alphabet. The Arabic alphabet consists of 28 letters.⁶ Of these letters only one is considered a pure vowel (alif) with the value of a long á.⁷ Two additional letters should be mentioned in this context: The yá' and the wáw, the so-called semivowels.

Depending on the word, these letters can be used as either vowels or consonants. The wáw can have the value of either a "w" or a long "ú" and the yá' can have the value of either a "y" or a long "í". Knowing which of the values to assign, depends on the basic meaning of the word or the grammar. This problem becomes evident only when one tries to read aloud or to transliterate the Arabic alphabet.

Furthermore, in the written language there is however a need for more vowel values than the three long ones: These values are the short "a", "i", and "u". They can be expressed in writing by the use of diacritic marks.⁸ Therefore, the written Arabic language consists of 28 (29) real letters and 5 diacritic marks. Although some of the grammatical endings are not expressed by real letters, they can be expressed by other diacritic marks.⁹ This adds three more symbols to the list of those needed to express written Arabic.

4 - There can be minor variations of the use of single terms and the use of vowels.

5 - The following list of problems concerning the written Arabic language in bibliographic databases only applies to natural language and not to controlled vocabulary, although some of these problems should be considered in the work of establishing future controlled vocabularies in Arabic.

6 - 29 if the hamza is considered a letter.

7 - I shall be using the Library of Congress Romanization Standard in this paper.

8 - The kashrah, fathah and dammah. Furthermore, it is possible to express the lack of vowels between two consonants by using the diacritic mark called sukun and the doubling of a consonant by using the diacritic mark shaddah.

9 - For the definite nunation the fathah, kashrah and dammah are used to express the three caseendings, for the indefinite caseendings three additional signs must be added the so-called tanwin.

As stated in note 6, the hamzah has a somewhat different position. The hamzah can be considered both a diacritic mark and a consonant: The hamzah can best be explained as a "stop-sound" and it must be attached to one of the real letters. One can consider the letters that can bear a hamzah as having a different value than the same letter not bearing a hamzah. The hamzah can also be considered a diacritic mark. If one considers the different combinations of hamzah and its different chairs as having separate values, this adds five more characters to the Arabic alphabet.

Finally, it is possible to express the long "á" in two more ways. An alif can be lengthened by adding a maddah, which also can be considered a diacritic mark. However, the maddah cannot be used in combination with any other letters. Secondly, the alif can be expressed (graphically) as the alif maqsúrah.

In terms of computers, 43 character values are needed in order to express the different values of the Arabic alphabet. Compared to the basic Latin alphabet, 26 values are needed if one ignores upper -

and lower case but 52 are required if one considers the upper - and lower case.

The second difference to note is the writing direction. Arabic is written from right to left, which is opposite of the Latin alphabet. In terms of databases, this creates a series of problems: As pointed out by Eilts¹⁰ records are stored in "logical order" i.e. the order in which they are typed. Every character is the order in which they are typed. Every character is thereby saved in accordance to "when it was typed" in the entire character string. In order to display the records it becomes the task of the display device to generate the right writing direction - that being left to right for some alphabets and left to right for other. As stated by Eilts "This is one of the most head-aching problems involved in multiscrptual computing"¹¹.

The third difference is the graphical expression of the letters. Arabic script is highly calligraphic (Kreipke, 1993) and most of the Arabic letters have 3 or 4 different graphical expressions¹². This means that they change their appearance according to their position, i.e. whether they are placed, initially, medially, finally

10 - John Eilts, personal communications, 24th. august 1995.

11 - John Eilts, personal communications, 24th. august 1995.

12 - 20 of the letters have 4 differcnt graphic expressions and 6 of them have three different graphic expressions.

in the word or isolated. However, they all maintain their phoneme value. This means that the Arabic alphabet consists of more than one hundred graphic expressions. To this number, one must add the different combination of hamzah and its chair and the fact that the combination of the letters lam-alif have a specific graphical expression.

The fourth problem discussed in this section concerns the numerals. There are two different graphic expressions for the numerals. The "Arabic" numerals (1,2,3,5, etc.) and the "Hindi" Numerals (٩ ٨ ٧ ٦ ٥ ٤ ٣ ٢ ١) . Both graphic expressions are used in Arab countries. The Arabic numerals, however, are mostly written or should be read from left to right that means opposite of the reading direction of words. "Mostly" refers to the fact that the same writing direction for numerals as for text can be observed¹³. So AD. 1995 can be expressed like 1995 or ٥٩٩١ or ١٩٩٥ in written Arabic.

The fifth difference is the lack of vowels in most written texts. In written

Arabic, only the long vowels are stated in writing¹⁴. The short vowels and other diacritic marks¹⁵ are mostly ignored. To a non-native reader, it seems difficult to understand how Arabic can be read without the presence of vowels. The technique consists of a mixture of guessing from the context and grammar and recognizing the graphic expressions of the words. Only when it is essential to understand the very specific meaning of the text (like in the Qur'an) the vowels (or diacritics) are stated.

The sixth difference is the basic structure of the Arabic language: the so-called roots and patterns. A very large percentage of the Arabic words consist of a three letter root (Al-Sadoun, 1989). Each root has a basic meaning. In theory; verbs, nouns, adjectives, and adverbs can be derived from each root¹⁶. All forms of words derived from the same root will to some extent be semantically related. To give an example, the basic meaning of the root **K-t-b** has to do with "writing"¹⁷. By changing the pattern, that means by

13 - This is only case when the hindi numerals are used.

14 - alif and the semi-vowels.

15 - The sukun and the shadda.

16 - Not all roots can be expressed in all forms or patterns and this rule only partially applies for names of persons and places and for foreign words - like "kumbútir" (i.e. computer).

17 - The basic meaning here is considered 3. person singularis, maskulinum, of the verb in the perfect tense .

changing the vowels or deleting the vowels between the consonants or doubling the root letters, by adding prefix, suffix or infix, one can varyate or change the meaning of the word within the framework of the basic meaning of the word. This is done in accordance with very strict grammatical rules. The verb *kataba* means "he wrote", the noun *kitáb* means "book", and the noun *maktabat* means library. To prefix *ma-* (in *maktabat*) has the basic meaning of "the place where....". In the example of *maktabat*, the combination of the basic meaning of the root (**k-t-b**: something that has to do with writing) and the prefix (*ma-*: something that has to do with the place) the meaning "Library" becomes evident. Another example is the fact that the doubling¹⁸ of the root letters "intensify" the meaning: *Kattaba* means "to keep up correspondence" or "to exchange letters". The system is fairly simple to use¹⁹. A native speaker is able to guess the meaning of words he does not know in a text from an often subconscious knowledge of the basic meaning of the roots, the patterns, the syntax, and the context.

This method works fine with humans, but is problematic when it comes to computers. The non-written vowels of the text are the chief problem. As the short vowels are not always written in Arabic, some of the written words in Arabic script have exactly the same appearance: to exemplify this the plural form of the word for book and the 3 person masculinum singularis perfect tense of the verb to write are both spelled *ktb*. However, when the vowels are added the words will be "*kataba*" or "*kutub*". The only way of knowing wich is by guessing from the context and grammar of the text. If the words are stated isolated, there is no way of knowing the exact meaning²⁰. And in the case of bibliographic databases which stores character stringe, and not meanings, this will necessarily affect the process of retrieval. This could lead to the conclusion that in terms of retrieving bibliographic records, the structures of these roots and patterns could be useful tools²¹, but in every language there are no rules without exceptions and this leads to the seventh problem.

The seventh problem concerns the

18 - Adding the shaddah.

19 - That is - if you know it.

20 - Sometimes an Arabic author might miht state one of the vowels in order to help the reader.

21 - This means if there was a way of using the semantic relationship between the words derived from the same roots for retrieval purposes.

problem of the so-called weak radicals. In the example of the root k-t-b this is a strong root. That means that the three letter combination of ktb is stable and will be recognized in exactly this form and order in all words and grammatical variations derived from this root. This is, however, not the case for all Arabic words. If one of the letters of the root is either a hamzah, one of the semivowels or alif, the root letters themselves can change during conjugation. A percentage of these weak radicals are considered irregular. Grammatical rules for the changing of the root values within the words exists, but they are many and indeed complicated.

The eighth problem relates to the case endings and most of the pronouns. Usually, the case endings are not written in a modern text²². This is due to the fact, that the case endings are stated as diacritic marks. They do, however, exist in very formal written Arabic. Only where the case ending affects the spelling of the word it will always be specified²³. Also the pronouns²⁴ are written as suffixes of the

words they refer to. If one wants to write "my book" the personal pronoun *i* must be attached to the word for book, where by this will be expressed like "kitábí" and her book will be "kitábha" etc.

The ninth problem concerns the orthography of Arabic. This problem arises from the fact that some particles in Arabic and the definite article are written without any separator. If one wants to express "he came and he left in his car" 7 separators (or spacebars) are used in English. In Arabic only 2 will be used: "yasul wayatrik wabisayyártihi"²⁵. In a transliterated version²⁶ in e.g. a bibliographic database separators will necessarily have to be added: "yasul wa-yatrik wa-bi-sayyaratihí". This also applies for several prepositions, for the definite article and finally for the two different ways of expressing "and" (wa and fa) and sometimes for the particle expressing vocative "yá".

A review of previous research

Scientific work in the field of establishing a character-set for Arabic script has been conducted as far back as in 1970

22 - The exception from this rule is in the indefinite accusative and in standard expressions.

23 - This is the case of the indefinite noun in accusative where an alif must be written after the last root-letter (and in very formal Arabic a diacritic mark will show that this alif is a case ending).

24 - except for the demonstrative pronoun.

25 - Notice that the subject is expressed implicitly in the verb through the conjugation.

26 - I shall use spacebar in accordance with how this is used in Arabic and "-" in order to express the division of the words and particles.

(Aman, 1984). It was not, however until 1982 that the first Arabic character-set for bibliographic work became available (Musa, 1986). The result was ASMO 449, a 7-bit character-set. This character-set still forms the basis of all the character-sets developed afterwards, and is considered the "Arabic ACSII". The next step came in the development of ASMO 708 (Ashoor, 1989) which is an 8-bit character-set. Despite these advances, the practical work in Japanese libraries proved, that to the East-Asian languages, the 7- and 8-bit character sets were not large enough. And in 1986 Sakai (Sakai, 1986) reported on efforts to add the Arabic characters to a 16-bit character-set. The problems concerning computer representation of all the scripts of the world led to different institutions working simultaneously on establishing character-sets for all the scripts of the world. These were ISO and the Unicode Consortium (Peruginelli, 1992). Fortunately, the two different standards merged, and in 1993 (Ksar, 1993) the first version of Unicode was released. Unicode contains all the values that are needed to work with the Arabic script on computer and is acknowledged as the international character-set standard. This standard, in theory, allows for an exchange of bibliographic records from different hardware

and software platforms and thereby solves the first problem mentioned concerning the Arabic alphabet.

Although Unicode resolves the character-set problems, the second problem concerning the writing direction to a certain extent remains. An article from 1980 (Madkour, 1980) reports on work done in Egypt in order to "Arabize" CDS/ISIS, but states that the problem concerning the writing direction had not yet been solved. In 1986 Musa (Musa, 1986) writes that most of the so-called arabized systems still do not have the correct writing direction, and that most of the arabized systems can only handle the opposite direction during the process of typing. After storing and retrieving, the graphic expression becomes the opposite of the intended. The problem still seems to be present in 1988 (Anees, 1988). As stated above, the problem is not a problem of the saving of records, the problem attaches to the display devices. Thereby it does not affect the command systems but the display device of the application. For a full description of these problems, Aliprand's two excellent articles describe how RLG solved the display problems for Arabic script (Aliprand, 1992a; Aliprand, 1992b). Furthermore, the solution chosen by RLG for RLIN solve the second problem concern-

ing the writing direction and thereby the display of bibliographic data in Arabic vernacular script.

The third problem concerning the graphic expression of the Arabic script still causes problems in some systems today. The problem, has to some extent, been solved through the so-called "Hydriyya method" described by Aman (Aman, 1984). It means that the basic form of the letters (i.e. the graphic expression of the letter in its isolated form) is stored as a code value. An interface generates the correct graphical expression - whereby the generating of the graphical expression is not actually stored, but formed by the computer when used. This method solved the problem of the need for more than a hundred different keys for Arabic script and made it possible to work in Arabic script with the correct graphical expression within the framework of a 7-bit code page. When I state that this still causes problems, it is due to the fact that the system named TINlib, produced by IME and sold to several libraries for its ability to handle Arabic script in accor-

dance with Unicode, is still not able to generate the correct graphical expression of the Arabic vernacular script²⁷.

Very little work has been done in order to handle the fifth and the sixth problem. These are concerned with the lack of explicit statement of vowels in modern written Arabic and how to handle this in a bibliographic environment. The most recent effort which analyzes these were made by Aliprand (Aliprand, 1992b) who explains how the diacritic marks of the Arabic script can be considered non-spacing characters. And Al-Sadoun's (Al-Sadoun, 1989) article presents the possibility of using the structure of root and pattern for compression of records in Arabic script.

For the ninth problem *khurshid*, (Khurshid, 1992) describes the awareness of the problem regarding the definite article which is written as part of the word it attaches to²⁸. In order to solve this problem, work has been done in Saudi Arabic to make lists of all the Arabic words, that starts with the letter combination *alif-lam* (the same combination as the definite article) and to make additional programming

27 - Unfortunately this is not documented in the literature and the information relies on conversations with Mr. Jean Smith from Library of Congress' regional office in Cairo during October 1994, Mrs. Benedikte Krag Schwarz from the Immigrants' Library in Denmark and from an e-mail received from Mr. Anton R. Pierce dated 15. of March 1995.

28 - He does not mention the problems concerning the particles.

for their database so that the retrieving system will ignore the letter combination alif-lam when written initially except for the words in the list.

As far as I am aware, no work has been done in order to solve the problems mentioned as the eighth and the fourth problem in this paper.

Analysis

In summary, the unsolved problems concerning records in Arabic script in bibliographic databases, are the lack of vowels (the fifth problem) which leads to the fact that there is no way of knowing the exact meaning of many words in Arabic if seen isolated (the sixth problem)²⁹. Furthermore, no one has yet solved the problems concerning the case endings and the pronouns written as a part of the word they refer to (the eighth problem). Neither the problem concerning the orthography of Arabic (the ninth problem), nor the problem concerning the numerals (the fourth problem) have been resolved. Some of these problems can be solved relatively easily, they only need an organization like IFLA to make a recommendation or to make the decision, and some of them can be solved locally as well. Others (the fifth, the sixth and the

seventh problem) need careful consideration and intensive research. In order to deal with the above stated remaining six problems, I shall start with the last as it seems to be the easiest one to solve (the fourth).

There are several ways of solving the fourth problem concerning numeric orthography. The problem becomes evident if an end-user wishes to conduct a search in natural language for literature in Arabic from all the Arab countries about a specific event. An example can be found in the search statement "The war in 1973". If the year-statement is supposed to be used for the search profile, should the search statement be "1973", or ١٩٧٣ or ١٩٧٣ or all of them?

Taking the AACR II into practice at this point would involve all three possible ways of expressing 1973; as the main rule of AACR II states, the information of the book should always form the basis of the bibliographic record. Typing the year statement in any other way than stated on the title page of the book is an inconsistency with this basic principle for cataloging. Therefore, the best way of solving this problem would be to add scripts to the existing bibliographic database sys-

²⁹ - I shall, for a short while, ignore the problem of the weak radicals but will return to it later in this paper.

tems. These scripts would first of all make the code values of the Arabic and the Hindi numerals equal in the searching facilities, and would mark records where the numerals are written from right to left, so that the underlying system would also retrieve these records for a search statement like 1973 or ١٩٧٣.

For the problem concerning the orthography of Arabic (the ninth problem) we are presently working on a solution at Odense University. We have not applied the solution presented by Khurshid (Khurshid, 1992). This solution was not chosen because it only deals with the definite article, and it also involves a very labor intensive work concerning the development of lists of words. Instead we developed a specific structure for the MARC-fields for the Arabic records. The basic idea is to place a separator between the meaning bearing words and the particles or the definite article during the initial cataloging of all the fields in natural (Arabic) language. We developed two scripts. The first script deletes the separator and generates the right form of the word and its particles and / or the definite article. The second script deletes all char-

acters prior to the separator. Thereby, we generate two versions of all natural language fields for each Arabic record. The first field include the second excludes them. For the searching facilities, we allow the end-user to choose whether to conduct the search with or without the particles or the definite article and this thereby avoids initial truncation for every search³⁰. Whether this proves to be the best solution is yet unclear, as the work is still in its initial phase.

As for the case endings and the personal pronouns (the eighth problem), we have not added a similar procedure. First of all because our database are an OPAC and it does not include abstracts. Therefore, the natural language statements in our database is mainly title information and / or cooperation information if not just author names. Samples of these information show, that the basic grammatical and syntactical structure of these statements are fairly simple and that case endings or personal pronouns only seldom occur. We have therefore chosen to consider the problem of personal pronouns and case endings in the same way as case-endings are handled for Indo-

30 - As meaning bearing words in Arabic also can be generated by adding prefixes to the root - an initial truncation will make the search profile too broad. To give an example: To search for books containing the root ktb in the title will retrieve books containing the word for library, office, typewriter etc. too.

European languages like German, Greek etc. This means, that the end-user will need to use truncation in order to strip the case-endings and / or personal pronouns from the search profile.

As for the problems that relates to the lack of vowels (the fifth problem), the structure of root and pattern and the weak radicals research needs to be done. These problems raise a lot of questions to be dealt with by future research.

First of all, work needs to be done in order to analyze the recall and precision in the existing databases containing records in Arabic script. This research should clarify whether the present systems give the same degree of precision and recall for natural language queries in Arabic as they do for English or Indo-European languages. Secondly, research should be conducted that analyzes the possibilities of improving recall and precision for natural language queries in Arabic. This research should involve a closer look at the

root and pattern structure of Arabic and an investigation of whether this structure can be used to improve the searching facilities.

From an overall point of view, using the root and pattern structures for queries should improve recall³¹, but will it thereby also effect the precision, and if it does, will it result in higher precision alongside with higher recall? Or a lower precision? Whatever the results show, we must also clarify how to deal with the weak radicals³², and how to deal with the lack of vowels³³: should we add them and thereby compromise the AACR II? These are some of the questions that eagerly await an answer.

Conclusion

Research in the use of Arabic script in a computer environment has come very far in course of the last two decades. It seems as if the basic problems concerning representation of the letters, the writing direction and the storing direction has been

31 - A search statement for the k-t-b root should retrieve titles containing all variants of words derived from this root. That means several semantically related terms, the plurals, the grammatical variations etc. It is yet unclear whether or not this is an advantage?

32 - The question is: can scripts be made for the weak radicals that enables the database to identify the basic root of the words?

33 - If it turns out that using the root - and pattern structure in search will result in a lower precision - how do we deal with the fact that due to the lack of vowel statements in modern written Arabic makes a lot of Arabic words look the same, because they consist of the same characterstring although they have different meanings كـب can be both kataba meaning "he wrote" and kutub meaning "books").

solved. However, for the specific use of Arabic script in bibliographic databases, i.e. for the retrieval process, a lot of research needs to be done. The basic question for the future research is how well or how poorly do the existing database systems handle the retrieval process. This could be measured by recall and precision and be compared to the similar figures found in studies on English databases. My thesis in this paper is that the basic structure of Arabic affects the retrieval process so severely, that we will have to think of an alternative retrieval process for Arabic material especially if we expect end-users to retrieve the material they need.

References

- Al-Sadoun, Sabah S. Al-Fedaghi and Humoud B. (1989): "Morphological Compression of arabic Text" (Information Processing & Management, Vol. 1990, No. vol. 26) p. 303 - 316
- Aliprand, Joan M. (1992a): "Arabic Script on RLIN" (Library hitech, Vol. 1992, No. vol. 10) p. 59 - 80.
- Aliprand, Joan M. (1992b): "Nonroman Scripts in a Bibliographic Enviroment" (Information Technology in Libraries, Vol. 1992, No. 10 (2)) p. 105 - 117
- Aman, Mohanned M. (1984): "Use of Arabic in Computerized Information Interchange" (Journal of American Society for Information Science, Vol. 1984, No. 35(4)) p. 204 - 210
- Anees, Munawar A. (1988): "Computers - Writing The Right Way?" (Pakistan Library Bulletin, Vol. 1988, No. vol 19/1) p. 1-9
- Ashoor, Mohammad Saleh (1989): "Arabization of Automated Library Systems in the Arab World: Need for Compatibility" (Libri, Vol. 1989, No. vol. 39 nr. 4) p. 294 - 302
- Beauvais, Francois (1984): "MINISIS á l'institut de Monde Arabe" (Documentaliste, Vol. 1984, No. 21) p. 150
- Khurshid, Zahiruddin (1987): "DOBIS / LIBIS Acquisition Subsystem in Operation at King Fahd University of Petroleum and Minerals" (Library Acquisitions: Practice & Theory, Vol. 1987, No. vol. 11) p. 325 - 334
- Khurshid, Zahiruddin (1992): "Arabic Online Catalog" (Information Technology and Libraries, Vol. 1992, No. September) P. 244 - 251
- Kreipk, Helle (1993): "Islamisk kalligrafi og arabisk skrift" (Mellemøstinformation, Månedsoversigt, Vol. 10, No.5) p. 10 - 14
- Ksar, Michael Y. (1993): "Untying Tongues" (Consensus, Vol. summer 1993, No. 4) p. 14 - 17

Madkour, M.A.K. (1980): "Information processing and retrieval in Arab countries: traditional approaches and modern potentials" (Unesco journal of Information Science, Librarianship and Archives Administration, Vol. 1980, No. 2) p. 97 - 104

Musa, F.A (1986): "A system for processing Bilingual Arabic / English text (Journal of the American Society for Information Science, Vol. 37, No. 5) p. 288 - 293

Panzer, Cecile (1992): "ALEPH - A multiscriptual, multilingual library management system" (Conference paper of the 3rd International Conference on Multilingual Computing, University of Durham, Center for Middle Eastern and Islamic Studies: 5,) p.5

Peruginelli, Susanna (1992): "Character sets: towards a standard solution" (Program, Vol. 1992, No. vol. 26 nr.3) p.215 - 223

Sakai, Yasushi (1986): "An experimental system for creating and managing Arabic Bibliographic Database - a Step Toward Effective International Information Exchange" (Libri, Vol. 1986, No. vol. 36, no. 4) p. 259 - 275

Wellisch, Hans Hanan (1976): "Script Conversion Practices in the World's Libraries" (Int.Lib. Rev, Vol. 1976, No. 8) p. 55 - 84

Wien, Charlotte (1994): Arabic Books in the Danish Research Library System. Paper presented at the ICEMCO, London 7 - 9 april.

