AraMosaic

The Arabic Mosaic Browser

- Introduction
- Getting/Installing AraMosaic
- Features
- Bugs/limitations
- Using XLANGBOX-ARA to create Arabic HTML files for AraMosaic:
- Technical Aspects
- Comments

Introduction

LangBox International is specialized in the Arabic support for UNIX Operating systems and applications. The LangBox team has been involved in several projects related to Arabization with constructors such as Silicon Graphics or SunSoft. In June 1996, after having seen several complains about the lack of Arabic Web Browser from our customer and on the ITISALAT mailing list, LangBox decides to investigate in the domain of Arabic Web support on UNIX platform.

The only solution we found for UNIX is the well-known PMosaic product and its Trilingual support (English/Persian/Arabic), which unfortunately does not support the ISO 8859-6 encoding codeset.

In order to contribute to the Arabic standard support on Internet, LangBox International technical team and its management has decided to study the Arabization of NCSA Mosaic using the XLANGBOX-ARA development package and to offer the result of this job to the UNIX Arabic User community. The experience of LangBox International in the Arabization process of applications and the knowledge of all its related issues has resulted in the delivery of the version "1.0" of AraMosaic.

Later in February/March 1997, the version "1.1" has been issued again by the team in order to fix several bugs and also to add a new codeset handling menu.
The Arabic codeset problem is the following: the UNIX world has adopted ISO 8859-6 for Arabic data encoding and LangBox International development on UNIX platforms goes in this way, but the fact is that while surfing on Arabic text Web servers, we found that several well-designed sites propose documents in two or three encoding codesets, generally MAC (i.e ISO 8859-6 compliant), MS Arabic (for Arabic Windows) and ISIRI. Unfortunatly, many of these sites just host pages in MS Arabic, just because they have been developed under a Windows PC platform. These different encodings are really a problem for the Arabic-Countries-Wide Web and it seems that Microsoft, even by having selected the ISO familly encoding for others languages (Cyrillic, Greek, Hebrew, ...) wants to keep its difference for the Arabic. Due to this, accessing an Arabic text page is always a problem and most Web designers sometimes just prefer to put Arabic text into GIF images, which is not efficient for indexing or keyword searching. Also, having several Arabic codeset causes problems when doing a keyword search on Web search engine: The user has to search keyword in ISO and then make a second search in MS to be exhaustive.
As long as Microsoft will keep this difference (which should not be due to technical reasons by the way) or as long as the UNICODE encoding will not be THE de-facto codeset over the Web, the Arabic Web will have to juggle with this different encoding problem.
Note: Latest News 03/20/97: It seems that Microsoft has decided finally to support ISO 8859-6 under the Arabic IExplorer 3.0x. So ISO becomes the common codeset for all Web browsers supporting Arabic. **We strongly recommend to develop Web pages using the ISO 8859-6 codeset**, at least before the coming of UNICODE.

In order to try to live with this political, marketing or whatever problem, the new menu of AraMosaic allows to select Arabic HTML documents stored in either ISO 8859-6, MS CP1256 and ISIRI 3342. Also an "Auto-Mode" flag should help to automaticaly detect the codeset between ISO 8859-6 or CP1256.

However, the internal AraMosaic codeset is **always** ISO 8859-6 and this new codeset support just allows to display or to print existing pages. Any string search, cut and Paste of text, file saving... will be done using ISO 8859-6.

This is a quick way to convert existing CP 1256 Web pages in ISO 8859-6 for the Web standardization.

Just load a MS CP1256 Web page, select "Save as" from the "File" menu and save the document under the ISO 8859-6 new filename document.

On April 1st 1998, the face of the Web browsing has changed : Netscape makes public access to its navigator source code under the Mozilla.org site. The X Mosaic source code is then not the only public web browser source code available. Then, some LangBox technical staff members are contributing on their free time to the Mozilla Language Enabling Feature - Arabic/Hebrew (Bi-Di) language Enabling project in order to promote the Arabic/Hebrew support on the Internet with the Mozilla team.

For more info on the global Mozilla Language Enabling project, click here

AraMosaic is an enhanced NCSA Mosaic 2.7b4 Unix/X11 WWW browser supporting Arabic and English text. like PMosaic, AraMosaic is considered derivative work, and its distribution and use are subject to terms set forth by Board of Trustees of the University of Illinois who have ownership of NCSA Mosaic. Press here to read copyright.

AraMosaic supports bilingual English/Arabic HTML documents sent from WWW servers to browsers using the standard HTTP protocol. The documentation/use of AraMosaic presume that you are already familiar with the WWW and NSCA Mosaic use. The actual basic codeset for Arabic HTML documents read and displayed by AraMosaic is ISO 8859-6. AraMosaic upon receiving bilingual hypertext will properly layout the text and images on the screen. WWW browsers which lack the ability to display Arabic upon receiving such a document will either display 8bits European characters. You can see a sample screen session by clicking here.

AraMosaic has been enhanced using the XLANGBOX-ARA Development environment. This version includes only the HTML page localization in Arabic, but menu, help messages or input area widgets (like "Find in Current"

menu) might be also easily localized by using XLANGBOX-ARA Arabic Motif library. Also, this version might not cover all Arabic language specific problems, but tries to fix major of them:

- Right to Left presentation of the HTML page, with scroll bar
- Arabic text context analysis for text shaping, with customization of some parameters (Latin/Hindi digits, Arabic diacritics on or off, Data processing mode or Word processing mode, Neutral characters handling in Right to Left mode,...)
- Text selection for Cut and Paste actions.
- Postscript printing of Arabic document.

AraMosaic is only available for Unix/X11 platforms at this time, however here also LangBox International is willing to provide solutions for PC/Windows and Mac in the future.

Getting/Installing AraMosaic

AraMosaic is available via anonymous ftp on the following sites:

- **ftp://ftp.langbox.com/pub/langbox/AraMosaic**
- **ftp://leb.net/pub/langbox/AraMosaic**

NEW| You can subscribe to the AraMosaic Update Registration Form to be automaticaly informed by e-mail of any change, new version, bug fixes...

AraMosaic is provided in binary form for the following systems:

- SGI 5.2/5.3/6.2 (AraMosaic.sgi.tar.gz)
- Sun Solaris 2.4/2.5(AraMosaic.solaris.tar.gz)
- Sun Solaris X86 2.6(AraMosaic.solarisX86.tar.gz) NEW
- SunOS 4.1.3/X11/OpenWindows (AraMosaic.sunos.tar.gz)
- Linux 2.x.x/Motif 2. (AraMosaic.linux_MotifDynam.tar.gz)
- Linux 2.x.x/no Motif (AraMosaic.linux_MotifStatic.tar.gz)
- DEC Alpha OSF1 3.2 (AraMosaic.alpha-dec-osf32.tar.gz) NEW

You must download the file corresponding to your Operating system with FTP as well as the **README.FIRST** file which details the installation process.

Then, the steps to install AraMosaic are the following:

- su
- mkdir /usr/local/AraMosaic
- cd /usr/local/AraMosaic
- gunzip -c AraMosaic.xxx.tar.gz | tar xvf -
- sh install.sh

AraMosaic has added Arabic fonts to your font server. Also, included for SunOs X11 systems, a XKeysymDB file allows to resolve Motif key bindings if warning messages result upon execution. See NCSA Mosaic FAQ for more info. To test for correct installation you may view the test file provided.

If you just typed "aramosaic HTML/AraMosaic-sample.html" or click here (only if you run AraMosaic), you will see the document in Arabic. Here is a sample screen output image of an AraMosaic session. See NCSA Mosaic documentation on how to use the Mosaic Web browser itself.

If upon execution, three warning messages are displayed:

```
Warning: Could not open font "-lbi-nashki-r-...-iso8859-6". Using fixed instead.
```

```
Warning: Could not open font "-lbi-nashki-r-...-iso8859-6". Using fixed instead.

Warning: Could not open font "-lbi-nashki-r-...-iso8859-6". Using fixed instead.
```

this means that the fonts were not installed correctly and you will see European characters instead of Arabic. Check your installation and the install.sh script file. You can check the Arabic fonts availability by running:

```
xlsfonts | grep iso8859-6
```

In the worse case, you must run manually the command:

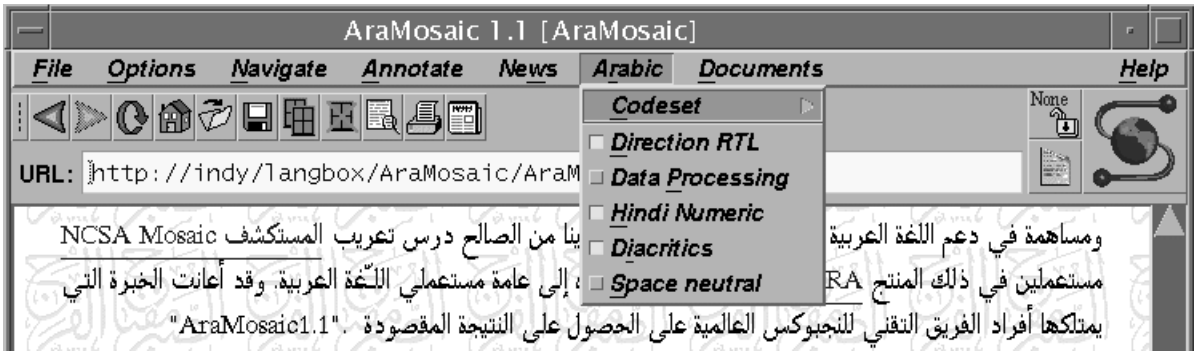
```
xset +fp /usr/local/AraMosaic/fonts
```

You can then check/see ISO 8859-6 Web pages on the WWW. Like NCSA Mosaic, this assumes direct access to Internet from your station.

We are trying to list some ISO 8859-6 Web sites on our Server, please check.

Features

New menus have been added to NCSA Mosaic 2.7b4. They are:



Arabic: A New Popup Menu has been added to the main menubar in order to allow specific Arabic language handling:

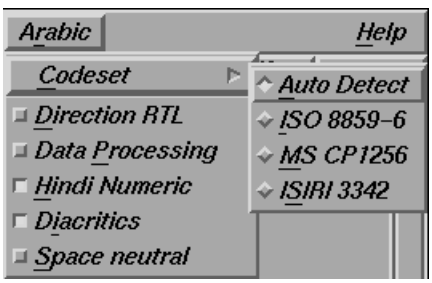
- Arabic Codeset: Select the Arabic data encoding codeset.
- Direction RTL: Toggle Latin and Arabic Global Writing Direction (Right to left and Left to right)
- Data processing mode: Toggle Word processing and Data processing mode.
- Hindi Numeric: Toggle Arabic and Hindi digit shapes.
- Diacritics mode: Enable or disable Arabic Diacritic management.
- Neutral Space mode: Set or Unset English space as a neutral character

The detailed meaning of these toggle is the following (you can find more detailed information on the XLANGBOX-ARA documentation):

- *The Codeset selection*

  This new menu of AraMosaic allows to select the Arabic HTML document encoding in either

  - ISO 8859-6,
  - MS CP1256 and
  - ISIRI 3342.



In fact the AraMosaic always support an ISO 8859-6 Arabic Context analysis engine, but Arabic data are converted on the fly by activating Codeset conversion routines.
The "Auto-Detect Mode" flag tries to automaticaly detect the codeset between ISO 8859-6 or CP1256 by analysing the content of the Arabic text itself, but this automatic detection is really efficient on long text (generaly more than one line). If you know the encoding of an Arabic document you can force it between ISO 8859-6, MS CP1256 and ISIRI 3342.
Also, since AraMosaic is based on the Arabic ISO 8859-6 context analysis engine of XLANGBOX-ARA, the Farsi ISIRI codeset selection allows to displays only Arabic document encoded using this Farsi codeset. Pure Farsi data are not supported and are stripped due to the lack of the Farsi characters in the AraMosaic fonts.

- *The Global Writing direction*

  Contrary to Latin-based languages, Arabic text is written from right to left. Because of this fundamental difference in writing direction, AraMosaic allows two kinds of sessions:

  The Latin (left to right or L2R) type session where the initial cursor position is located at the leftmost position of the text widget, and text is written from left to right.

  The Arabic (right to left or R2L) type session where the initial cursor position is located on the rightmost position of the text widget, and text is written from right to left.

- *Arabic Data Storage and Display*

  AraMosaic allows the user to work with two different Arabic codesets internally:

  - ASMO 708 (ISO 8859-6): Digits are always 7 bit encoded.
  - ASMO 449+: Digits can be 7 bit or 8 bit encoded.

- *Arabic Numerals*

  Arabic digits, or Numerals, are written from left to right, as in Latin languages. Arabic digit may be displayed in either Hindi or Arabic digits depending on the choice of the user.

- *Diacritic or "Tashkil" generation*

  AraMosaic manages and displays the vocalization characters witch are supported by ISO 8859-6 and ASMO 449+ codesets. They are the following :

  - The Shadda
  - The Sunkun
  - The Fatha
  - The Damma
  - The Kasra
  - The Fathatan
  - The Dammatan
  - The Kasratan.

- *Handling of Neutral characters and Spaces*

  Arabic and Latin characters conflict in the direction of the display. When writing Arabic in an English line, characters are pushed on the line as they come from the keyboard or from a file. The reverse effect happens when entering an English character in an Arabic line.

  The user may define Latin Space as neutral characters which will follow the global writing direction despite its language value. This feature is useful when displaying Latin tabulated text in Right to Left mode. The typical case is when browsing directories contents using URLs beginning with ``ftp://'', for example, ftp://www.langbox.com/pub/langbox

Options/Fonts/Arabic: A new Arabic font allow to select Arabic font or to switch back to regular

Latin/European ones.

AraMosaic comes with two Arabic in three sizes;

- Proportional width font (like Times for Latin), sizes: 12 14 24
- Fixed width font (like Courier for Latin), sizes: 12 14 18

These fonts are sufficient to display and read Arabic HTML document. However, additional fonts might be available under XLANGBOX-ARA package. The AraMosaic Arabic fonts are installed automatically during the AraMosaic installation process. They are added to your X font Server.

Mosaic HTML widget display has been enhanced to support:

Right alignment of the HTML page

In order to present correctly all HTML element, the whole page is Right aligned (Image, text, bullets...)

The Cut'n Paste feature

The standard Cut'n Paste feature is transparent and is compatible with the X Server Cut'n Paste buffer. User may cut an Arabic string from an AraMosaic session and paste it in an (8 bit) editor.

Although, very useful and essential, this feature is technically not easy to implement for BiDi languages. We noticed for example that PMosaic is not handling it at all and some other Windows Arabic Browser have limited this feature to the selection of full entire line only to avoid complication.

Under AraMosaic, the user can select a text from one character to an other, like for the standard English version and the selection is done on the Logical order (i.e. the order of text input). This can give unusual results when the user tries to select mixed Latin/Arabic text in one selection. The highlighted area might be split into one, two or three different visual sections. In any case, the internal selection buffer contains a consecutive logical buffer. According our XLANGBOX-ARA support experience, users are familiar with this feature after 15 minutes of use (and after all, this is also the solution adopted under Microsoft Arabic Windows).

Horizontal scroll bar

Horizontal scroll bar is set automatically aligned on the Right side when displaying Right to left orientation, and on the Left side when displaying Left to Right orientation.

Print of Latin/Arabic documents is supported in Postscript.

Postcript printing of Arabic document is supported under AraMosaic. When the Arabic HTML page is loaded, select Print menu and Postscript format. A Postscript document is built and sent to the printer through the AraMosaic printer command. This feature presumes that the user has already a Postscript printer correctly set up on his system.

Bugs/Limitations

Please report bugs to us first, NCSA Mosaic 2.7b4 is quite stable and any core dumps are mostly likely due to our additions. If the bug is confirmed not to be from our areas, we shall inform the already too busy NCSA team.

Known bugs/limitations:

- Automatic right alignment detection may not correctly in some situations, mainly in line of English only text that includes several type of element. (for example: LatinText1, LatinLink, LatinText2, in RTL, elements should be permuted and not easy to read due to an individual presentation from Right to left)
- Table don't support Anchor data (but this is a NCSA Mosaic 2.7b4 limitation).

Using XLANGBOX-ARA to create Arabic HTML files for AraMosaic

Creating Arabic hypertext files which AraMosaic can display is quite easy. Arabic HTML is no different than standard HTML. Simple begin by creating ISO 8859-6 text section encoding using any of your favorite tools. Since XLANGBOX-ARA encoding uses this character codeset, users can use axmedit to edit/add Arabic text in HTML document.

You can also uses any other Editor from the market that support this codeset (this is the case for the Arabic Mac tools).

Also, by the merge of using the Arabic Motif library of either ALM under Silicon Graphics IRIX or XLANGBOX-ARA under Sun Solaris, AraMosaic can handled and display Arabic menu labels as well as bilingual Input areas in the HTML document. It become possible to search for an Arabic string within an HTML document for example or to fill a CGI form with Arabic data.

Technical Aspects

AraMosaic Beta 1.0 supports as its default encoding ISO 8859-6, the current ISO character set for Arabic encoding. The AraMosaic 1.1 supports in addition Codeset conversion from MS CP1256 and ISIRI 3342. These codesets are "8 bits codesets". The lower 128 characters reflect 7 bit ASCII, and the upper 128 characters are used to represent Arabic. If you are Arabic User, you might already be familiar with these codesets. This limits HTML documents to Bilingual documents and in any cases, but this the case for all 8 bits codeset applications. This may change in the future if the default character set might be UNICODE (ISO10646) and AraMosaic will only display Arabic or Latin if the recognized characters are encoded in the Arabic code page.

We were first trying to reach the "Transparency" use of Mosaic, and that why we haven't modified or extended the HTML language with some additional markup. However, we are following all discussions done on the Bilingual/Multilingual WWW support, as well as other similar work such as PMosaic and we are aware of the need to extend also the HTML to include new markup such as Charset, Language, Direction... in order to complete AraMosaic. Providing a Web Browser BIDI extension should be closely linked with the extension of the HTML language in order to define additional features:

- Direction markup extensions (RTL/LTR) for markup such as <table>, <ul>, <il> or block <p> for example. This allows the ability to have both RTL and LTR sections within the same HTML document. (althrough the Horzontal scrolling should be disabled in such cases...)
- Charset encoding field (set to "ISO 8859-6" for Arabic) or <lang> markup allowing the activation of BI-DI support for Arabic and display of real Multilingual HTML pages (French, Cyrillic, Arabic...).
- The new font downloading markup, allowing the browser to download a TYPE 1 Postscript scaleable font by interpreting an image file embedded into a GIF file for example.

For Other Browsers, currently our ALM or XLANGBOX-ARA X11/Motif library allows users to display ISO 8859-6 HTML pages under Netscape Navigator, but the cursor pointing or selection feature cannot be handled by a

solution located only at the X11 level. The Main HTML widget window needs to be modified to support Right-to-Left orientation languages (i.e. display, cursor pointing, selection highlighting). The <select> <input>... widgets should be directly handled with the Arabic Motif library of ALM or XLANGBOX-ARA, since it seems that they are not Netscape built-in widgets, but OS libraries calls.

We are also examining the plug-in feature possibilities in order to realize this support from outside Netscape. However, handling this within the Netscape main HTML widget should be more efficient and elegant. In addition, such a Plug-In should handle all the HTML language elements, such as Java, JavaScript, HTML 3.2, Animated GIFs..., which not really the purpose of our contribution.

Comments

Feel free to send your comments, feedback, questions and reviews to aramosaic@langbox.com.

Last Time Modified : May 30, 1998
Copyright © 1998 LangBox International - All rights reserved