Standard for Arabic Inclusion and Support

Muhammad Alkarouri, Arabeyes Project `<contact (at) arabeyes dot org>`

Munzir Obeid, Arabeyes Project

$Revision: 1.3 $

This project aims to standardize the process of certifying Linux applications and distros with respect to Arabic support.

Table of Contents

License

Copyright (c) 2003, Arabeyes Project, Munzir Taha & Muhammad Alkarouri.

Background

While the use of Arabic in Linux is possible, the support in various Linux distributions is not straight

forward. This project/document was initiated to advance the Arabic support in Linux by recommending various Arabic supporting software and pushing for Arabic supporting default configurations in Linux distros.

Credits

To give credit where it is due, we must thank Nadim Shaikli and Mohammed Elzubeir for their technical support, administrational support and encouragement.

Feedback

The sources for this document vary from different mailing-lists, personal experiences and feedback from others on what needs to be addressed.

If you have any questions or suggestion, please do make them known on the 'doc' mailing list here: http://lists.arabeyes.org/mailman/listinfo/doc [http://lists.arabeyes.org/mailman/listinfo/doc]

Introduction

Standard for Arabic Inclusion and Support

The Standard for Arabic Inclusion and Support (Henceforth abbreviated as SAIS) is a group of specifications that should be complied with by a specific application or distribution on the GNU/Linux platform to be declared as 'Arabic Friendly'. To this end, the standard defines *Arabization Specification Levels* or *ASL* that are used for certifying an application or distro as supporting Arabic to a certain extent. There are currently two different sets of ASL, one for applications and another for distributions.

What is Arabization?

The word Arabize is born in 1883. - - Ar*ab*i*za*tion /"ar-&-b&-'zA-sh&n/ noun

1. a : to cause to acquire Arabic customs, manners, speech, or outlook b : to modify (a population) by intermarriage with Arabs

2. Arabicize(1)

In this context, a piece of software is said to be Arabized when it supports the Arabic language in reading and writing in addition to various locale-specific information.

Obviously, different applications need different aspects of Arabization, and some points above are more important than others. In particular, the best point upward for the support of the Arabic language is through UTF-8 and the openi18n initiative (http://www.openi18n.org).

Why we need standardization?

- Standardization is vital. lack of standardization prevents integration and accumulation of expertise, efforts, and results.

- lack of knowledge regarding what the Arabic community wants and what this community sees as partially Arabic supported and what is fully Arabic supported.

- Many problems in the support of Arabic are due to sheer ignorance from the side of developers about these problems. We can tell them now through this document to judge their works against these standards and the community can help giving weights or grades to the support level according to this standard.

- Developers should be aware of the context-sensitivity of the Arabic alphabet. We should tell them that they have to include "Tashkeel" marks as a basic requirement to get a context-free Arabic written word.

- Announcing our drafts/standards for things which is not standardized yet. Example: the mapping proposed by Muhammad Sameer of the Holy Quran marks in Unicode.

Topics for future standardization

There are many points that should be discussed within the community and that warrant future standardization. Examples are the treatment of different locales, identifying the minimum set of *tashkeel* necessary to preserve the context-free property, whether a software that does not support tashkeel will be regarded as supporting Arabic and to which extent.

A mechanism should be created within the Arabic speaking community for drafting such standard and submitting them to various standardization bodies and internationalization communities.

Arabization

Arabization Standard Levels

There are different Arabization Standard Levels defined. A group of ASL is defined for applications. Another group is defined for Distributions,

Software Arabization

Arabization of a certain piece of software involves the following points:

1. Correctly displaying and accepting as input UTF-8 characters representing Arabic letters. This includes complicated issues as BiDi/CTL, RTL vs. LTR paragraphs, etc.

2. Conforming to the current locale for date, currency, and digits (Arabic vs. Hindi).

3. The inclusion of an Arabic interface with Arabic menus,prompts,etc

4. Supporting other encodings (e.g. ISO-8859-6)

Arabization for Applications (SAIS - Applications)

Applications here are classified in the following categories, needed for the comfortable use of a desktop environment:

1. Base Libraries: like fribidi, ICU

2. Text user interface: consoles and shells: like mlterm, Akka

3. Programming Languages and programming environment: support for UTF-8

4. Graphical Desktop Environments: KDE, GNOME

A program is said to be 'adequately Arabized' if it is certified as supporting UTF-8 and locale conformance. It is said to be 'fully Arabized' if it is certified as supporting an Arabic interface. These terms refer to the levels ASL-Green and ASL-Blue below.

Arabization Standard Levels - Applications

SAIS makes use of a double terminology to indicate the level of Arabization standards. The first terminology makes use of stars, and scales applications on a one-star to five-star scale. A color-coded indicator can also be used, and should be beneficial for color-coding in tables, etc. The Arabization Standard Levels for Applications are:

Red (*)

An application is said to be *ASL-Red* if it needs Arabic support but does not have it yet. This serves as a 'null' level.

Orange (**)

An application is said to be *ASL-Orange* if it has wide character support with planned UTF-8 support, or that has UTF-8 support that is buggy, e.g. lacking Arabic shaping

Yellow (***)

An application that has full UTF-8 support as described in condition 1 of Software Arabization is said to be *ASL-Yellow*.

Green (****)

An application that fulfills the ASL-Yellow conditions and satisfies condition 2 of Software Arabization by supporting the Arabic locales is said to be *ASL-Green* or *adequately Arabized* in the ASL sense.

Blue (*****)

An application that satisfies the conditions for ASL-Green certification and, additionally, includes an Arabic UI as specified in condition 3 of Software Arabization is said to be *ASL-Green* or *fully Arabized* in the ASL sense.

Application Classification - Prerequisites

The following information are needed in application classification, both to specify the situation of the package itself and as its value for a distribution.

Name	is the name of the application/package.
Version	is the version subject to classification
Functionality	is the what the program does. All programs labeled with the same functionality can be substituted for one another without a significant loss of functionality.

Importance   is the degree of need of the program in a distro. The following choices can be made:

1. A must

2. Critical

3. Important

4. Nice to Have

5. Of Interest

Status   is the situation of the program with respect to Arabization. Can range from ASL-Red (for a program listed for its importance but needed for Arabization) to ASL-Blue.

Comments   Any particular comments related to the Arabization status of a package or its packaging.

Packaging   is the packaging form available for this application. Examples are .rpm .deb and .tgz. More than one is possible.

Application Certification Procedure

An application is submitted by one of its developers or anybody else who can step forward as an Application Arabization Maintainer for the specific application. The Application Arabization Maintainer should be responsible for testing the application in the version submitted and later ones.

The Application Arabization Maintainer should provide the information above mentioned. In particular, he should propose a certain *status*/level for the package. This status is verified by a second person in the Arabeyes community before it is submitted to be included in the applications table. The *importance* attribute should be submitted by the Application Arabization Maintainer as a suggestion. The actual importance will be awarded by the community.

Applications which are specifically developed with Arabization in mind, such as the Arabeyes internal projects, are assumed to be maintained by their own developers. The normal testing and bug squashing for such applications is sufficient for asserting their status.

A Sample Application Table

The following is a table that serves as an example for the application. *To be done: this table will be removed in favor of a separate real table in next revision*

Name	Version	Status	Functionality	Importance	Comments	Packaging
mlterm	2.7.0+	***	terminal emulator	Must have	minor bugs	rpm
fribidi	0.10.4+	***	BiDi support library	Must have	Shaping missing	rpm

Duali	0.1.1	****	spelling checker	Must have	functional, UI not polished	rpm,deb
Vim	6.2+	****	text based editor	Must have		rpm
Katoob	0.3.5	****	GUI based editor	critical		rpm
KASCT fonts	1.4	*****	fonts	critical		rpm,deb
Khotot fonts	1.0	*****	fonts	critical		rpm
fvwm	2.5.7+	****	Window Manager	important		rpm,deb
ArabTex	3.x?	****	Arabic typesetter	Must have		archive(?)

Arabization for Distributions (SAIS - Distributions)

Introduction

The certification procedure explained in this section aims to encourage distros to include the latest bug-free applications that support Arabic and to enourage them to apply proper default configurations to enable Arabic. On the other hand, the responsibility of advancing application compliance falls on the application developers.

Certification can take place only for a finished distribution. Beta and unstable distros cannot be certified, as their status can change very frequently.

The exclusion of a specific application due to a legal problem (e.g. being not GPL compatible) does not prevent a distro from getting a certain level of certification

Arabization Standard Levels - Distributions

There are three levels of Arabic support in distributions. They are the Silver Level, the Gold level and the Platinum level. The

Platinum

A distro is said to be supporting Arabic at the *Platinum level*, or *fully supporting Arabic*, if it supports all major variations of applications in the must-have, critical and important level and, in addition, having an installer and distro-specific applications supporting Arabic.

Gold

A distro is said to be supporting Arabic at the *Gold level* if the distribution includes at least one program that satisfies each functionality at the Must Have or Critical level in such a way that the program in

question is packaged with the default distribution and its default configuration is working with no tweaking, possibly with specific installation-time decisions made.

Silver

A distro is said to be supporting Arabic at the *Silver level* if the distribution is augmented with a single how-to that enhances it so a user can use the how-to to elevate the distro to a gold-equivalent one. The how-to can include:

1. The public place in the internet from where to download packages packaged for the distro to satisfy the must-have and critical dependencies.

2. Any configuration steps needed to make the above applications or ones included within the distro working.