

National Language (Tamil) Support in Oracle
An Oracle White paper / November 2004

Vasundhara V* & Nagarajan M[&]

* vasundhara.venkatasubramanian@oracle.com; [&]Nagarajan.muthukrishnan@oracle.com)

Oracle India Pvt Ltd.

9, Prestige Blue Chip, Level III, Floor III,
Hosur Road, Bangalore – 560029 India.

Table of Contents

INTRODUCTION.....	2
WHY LOCALIZE?.....	3
TARGET USERS	3
APPLICATIONS OF LOCALIZATIONS	3
ORACLE SERVER.....	3
LANGUAGE SUPPORT.....	4
TERRITORY SUPPORT.....	5
DATE AND TIME FORMATS.....	5
MONETARY AND NUMERIC FORMATS.....	5
CALENDARS.....	5
LINGUISTIC SUPPORT.....	5
CHARACTER SET SUPPORT.....	5
ORACLE TOOLS.....	6
PORTABLE JAVA COMPONENTS.....	6
JAVA INPUT METHOD ENGINE.....	6
DISCOVERER – A Business Intelligence tool.....	7
CONCLUSION.....	8
REFERENCES.....	8

Introduction

Globalization is the key to growth in this age. Inherent to the idea of globalization is the concept of localization. With the changing rules of business and enhanced role of Internet in, potentially, boosting the top line of the business, it is of paramount importance that businesses revamp their storefront to cater to global customers. At the very least, visitors to their online store, should be able to see the content in their own language, with the content dynamically transforming itself to cater to his own territory definitions, e.g., currency sign, date format etc. Just as global businesses need to connect in a 'local and friendly' manner, the local agencies also have come to tap the strengths of computing. To that effect, computing needs of communities too have to be provided for. Supporting a national language solves the dual purpose of global companies' need to have a local presence and local agencies/companies to be able to effectively use computing for their purposes.

Oracle products have always been designed with this very underlying principle; so that when one uses Oracle, one does not have to worry about how to make their business global. The Dravidian and Indo-Aryan languages, one of the prominent ones being Tamil, are vastly categorized as Indic languages. Since, Unicode is the base for storing all the data for the Indic languages, Oracle has the built-in support for Tamil. Unicode is a universal encoded character set that enables one to store information in any language, using a single character set. Unicode provides a unique code value for every character, regardless of the platform, program, or language. When Unicode is used in an Oracle database, it has the following advantages:

- Σ Simplifies character set conversion and linguistic sort functions
- Σ Improves performance compared with native multi-byte character sets
- Σ Supports the Unicode data-type based on the Unicode standard

While urban India is predominantly English speaking, the vast majority of Indian population who live in district towns and countryside are not. Indian constitution provides for 14 specified languages for Government work, English being one of them.

There have been initiatives by every state government to implement the state's native language for official communications. Property registrations, land records are predominantly made in native languages in rural India. So are the Tax notices, Government gazettes etc. Ability to provide utilities bills by government held institutions in the native language is growing in demand. Education has always been delivered and best understood in one's mother tongue. This is more important when offering adult education at the district and village levels.

Oracle provides the Database product to store data in the native language and additional tools encompassing the entire 3-tier architecture that facilitates complex operations on this data. Thus, Oracle provides an end-to-end solution for applications that need to interface in Tamil.

Why Localize?

The benefits of localization are

- ∑ The target audience for an application increases
- ∑ Develop once use anywhere
- ∑ Easier Web deployment

Target Users

Government - Central and State governments for e-Governance applications, Government departments, PSU's, Educational and Research Institutions etc.

Web Sites - Portals wanting to develop Indian language channels, Newspapers, magazines hosting online editions, Govt/ PSU's/ educational/ Institutional web sites etc.

Applications of localizations

There are endless possibilities to use localization. Some of the applications for which Oracle's solution could be used are:

E-governance applications	Customer Care & Billing
Census records	Land Records
Generic Records	Kiosks
Web Portal and Web Sites	Public Information Systems
Email	Chat
Online forms	Information archives
Wireless messaging	Data warehousing and data mining
Document publishing and printing	Web search engines
Business Applications	

Oracle Server

Oracle offers a unique capability that is unparalleled and solves a long-standing problem of true data processing capability in Indian languages. The solution is Unicode compliant and can be used with Oracle RDBMS for data entry, indexing, searching, sorting and generating reports. The data is universally portable and can be displayed using any Unicode compliant fonts and application software. Data and text produced using Oracle's solution can be converted to formats such as XML, HTML and WML and hence can be re-purposed for use in any environment. Since it uses Java as the backbone, applications developed using Oracle's technology can be deployed on any hardware/OS platform and the data produced by such applications will be universally exchangeable and portable.

Oracle server has the required enhancements to handle Tamil characters. Oracle Server's National Language Support (NLS) architecture allows one to store, process, and retrieve data

in native languages. It ensures that database utilities and error messages, sort order, date, time, monetary, numeric, and calendar conventions automatically adapt to the native language and locale.

Oracle RDBMS is implemented using three-tier architecture. The language-dependent operations are controlled by a number of parameters and environment variables on the client, the middle tier and the server. On the server, each session started on behalf of a client may run in the same or different locale, and have the same or different language requirements specified.

A database itself also has a set of session-independent NLS parameters (CHARSET parameters) specified at its creation time. Two of them are the database and the national character set. They specify the character set used to store data in the database. Other parameters, like language and territory, are used in the evaluation of CHECK constraints.

In the event that the client and server specify different character sets, Oracle will handle character set conversion of strings automatically.

As far as NLS (National Language Support) architecture is concerned, all applications, even those running on the same physical machine as the Oracle instance, are considered clients. For example, SQL*Plus started by the Unix user which owns Oracle software, from the Oracle Home in which RDBMS software is installed, and connecting to the database through an adapter by specifying the ORACLE_SID, is still considered a client and its behavior is ruled by client-side NLS parameters.

When a client application is started, it initializes its client NLS environment from environment settings. All NLS operations performed locally are executed using these settings.

When the application connects to a database, a session is created on the server. The new session initializes its NLS environment from NLS instance parameters specified in the initialization parameter file. These settings can be subsequently changed by an ALTER SESSION statement. The statement changes the session NLS environment only. It does not change the local client NLS environment. The session NLS settings are used in the processing of SQL and PL/SQL statements executed on the server.

LANGUAGE SUPPORT

Oracle allows users to store, process, and retrieve data in Hindi, Tamil, Kannada, Malayalam, Gujarati, Oriya, Punjabi, Telugu, Assamese and Marathi. Through the Unicode (UTF8) character set.

Additional support is available for a subset of the native languages, for which Oracle knows, for example, how to display dates using translated month names or how to sort text data according to cultural conventions.

Documents can be stored in multiple languages in CLOB, NCLOB, or BLOB data types and set up Oracle Text to enable content search for the documents.

TERRITORY SUPPORT

Oracle supports different cultural conventions, which are specific to a given geographical location. Local time, date, numeric and monetary conventions are handled. From Oracle 8.1.5 onwards, Oracle has been supporting “India” territory.

DATE AND TIME FORMATS

The world's various conventions for hour, day, month, and year can be handled in Indian language formats.

MONETARY AND NUMERIC FORMATS

Currency, credit, and debit symbols can be represented in Indian language format. Radix symbols and thousands separators can be defined by locales (Hindi_India.UTF8 in this case).

CALENDARS

Gregorian, Japanese Imperial, ROC Official, Thai Buddha, Persian, English Hijrah, and Arabic Hijrah are supported. The one used by the Indian locale is Gregorian.

LINGUISTIC SUPPORT

Oracle provides linguistic definitions for culturally accurate sorting and case conversion of Hindi, Tamil, Kannada, Malayalam, Gujarati, Oriya, Punjabi, Telugu, Assamese, Bengali and Marathi languages.

CHARACTER SET SUPPORT

Oracle supports a large number of single-byte, multi-byte, and fixed-width encoding schemes, which are based on national, international, and vendor-specific standards. The character set for Indian languages is UTF8.

Indian language specific data in the database can be stored in Unicode format. Oracle server has been changed to handle Indian language characters like month names, day names, and AM/PM/AD/BC string. These features can be used to store data in any of these languages.

Oracle Tools

Oracle provides Forms and Reports tools that can be used for data input and retrieval. The features that are provided by the localization of the Oracle tools are:

- Σ A set of pluggable java components, which could be used to develop applications in Indian languages using Oracle tools
- Σ The Keyboard layout and rendering are as per *Inscript* format.
- Σ A Java Input Method Engine for Indian language is also provided that could be used with JDK 1.3.x to input data in any of Java applications.
- Σ Unicode fonts pertaining to all the supported languages can be procured from font vendors

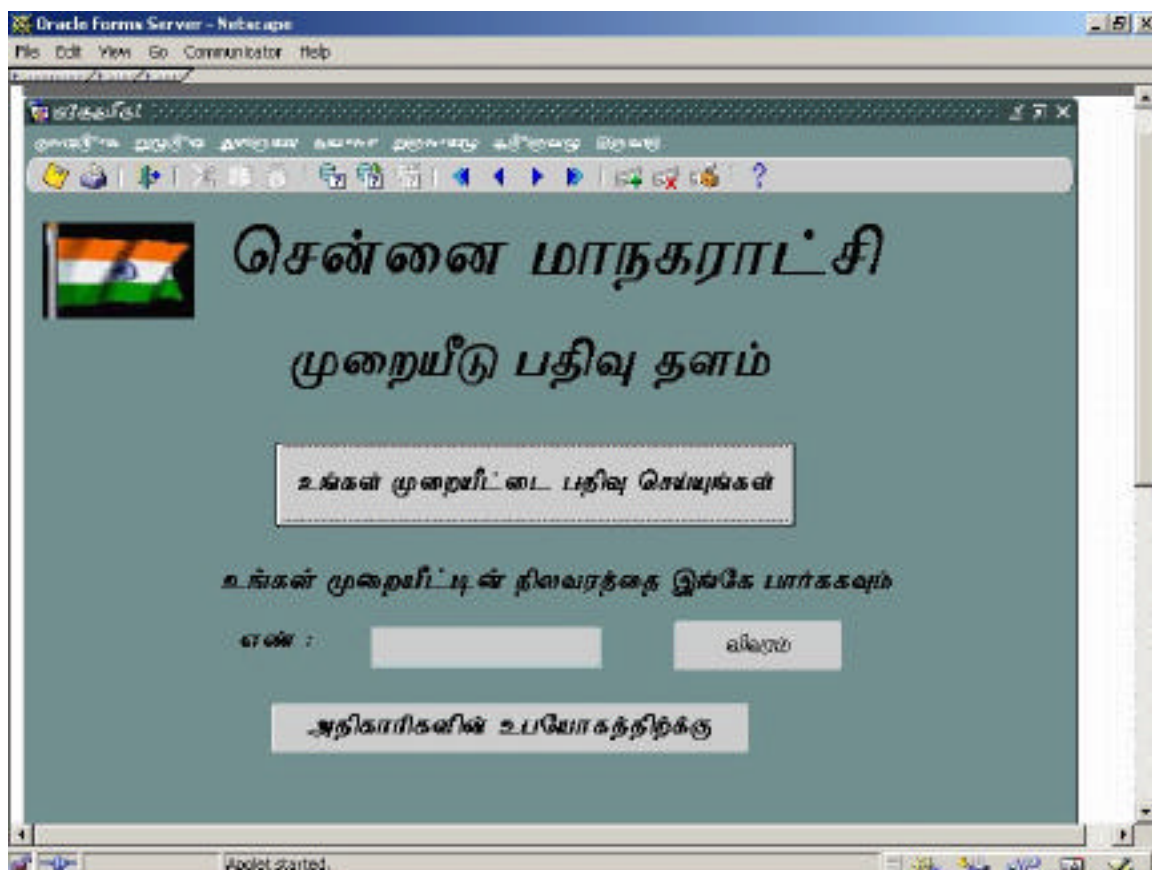
PORTABLE JAVA COMPONENTS

Oracle tools support Java client side user extensions through the use of standard JavaBeans and specialized Java components called Pluggable Java Components (PJC's). Using PJC's and JavaBeans it is possible for application developers to insert custom code to run inside the Forms applet and to tailor or customize the default operation of the default Forms Java components.

JAVA INPUT METHOD ENGINE

Input methods are software components that let the user enter text in ways other than simple typing on a keyboard. The input method framework in the Java 2 platform enables the collaboration between text editing components and input methods in entering text. Its service provider interface (SPI) for input method engines provides interfaces that enable the development of input methods in the Java programming language that can be used with any Java runtime environment; it also supports native input methods of the host platform. The Java Input Method Engine is developed for Indian language. The figure below contains a sample screen developed using Pluggable Java Components and Java Input Method Engine¹.

¹ The screen represents a demo that has been internally developed at Oracle. It is not a part of any live application.



DISCOVERER – A Business Intelligence tool

Oracle Discoverer, a key component of Oracle Application Server's integrated business intelligence solution, is an intuitive ad hoc query, reporting, analysis, and web publishing tool that allows business users at all levels of the organization to gain immediate access to information from data marts, data warehouses, and online transaction processing (OLTP) systems. Discoverer enables report builders and analysts to create, modify, and execute ad hoc queries and reports. Casual users can view and navigate through predefined reports, graphs and business views that hide the complexity of the underlying data structures being reported upon.

- Σ Oracle Discoverer 4i – Plus and Oracle Discoverer 9i – Plus have support for Tamil
- Σ Oracle Discoverer 4i – Viewer and Oracle Discoverer 9i – Viewer have support for displaying in Tamil over Internet Explorer and Mozilla browsers

The figure below shows a sample output from one of the Discoverer pages that have Indic language support.

 Assamese_অসমীয়া
 Bangla_বাংলা
 Gujarati_ગુજરાતી
 Hindi_हिन्दी
 Kannada_ಕನ್ನಡ
 Malayalam_മലയാളം
 Marathi_मराठी
 Oriya_ଓଡ଼ିଆ
 Punjabi_ਪੰਜਾਬੀ
 Tamil_தமிழ்
 Telugu_తెలుగు

Conclusion

Oracle offers a completely portable, maintainable and scalable solution that caters to any business requirement for Indian market. It enables developers to write software that can be deployed easily across a variety of world markets. This suite of tools will aid in storing, retrieving and processing Indian Language Data and also provide a means to input data in their native language in a way that conforms to the rules of their languages. The database is able to store multi-lingual data, which is universally exchangeable and portable.

References

<http://www.oracle.com>
<http://www.otn.oracle.com>
<http://www.oracle.com/global/in/languages/tamil.html>



National Language (Tamil) Support in Oracle

Nov 2004

Authors: Vasundhara V (vasundhara.venkatasubramanian@oracle.com)

Nagarajan M (Nagarajan.muthukrishnan@oracle.com)

Oracle India Pvt Ltd.

9, Prestige Blue Chip, Level III, Floor III,

Hosur Road, Bangalore – 560029 India.

Worldwide Inquiries:

Phone: +1.650.506.7000

Fax: +1.650.506.7200

www.oracle.com

Oracle is a registered trademark of Oracle Corporation. Various product and service names referenced herein may be trademarks of Oracle Corporation. All other product and service names mentioned may be trademarks of their respective owners.

Copyright © 2004 Oracle Corporation

All rights reserved.