11. Indian Language Support in Microsoft Products

Mr.Raveesh Gupta, Microsoft India, *E-mail: raveeshg@microsoft.com*

Indian Language support in Microsoft products started around 1999 with Windows 2000 Professional being the first official product to provide Indic support. It supported 2 scripts: Devanagari and Tamil; and 5 languages: Hindi, Konkani, Marathi, Sanskrit and Tamil. Windows XP saw more Indian scripts and languages being supported in MS products: Gujarati, Kannada, Punjabi and Telugu. Currently with Service Pack 2, Windows XP now supports 11 Indian Languages:

1. Bengali
2. Gujarati
3. Hindi
4. Kannada
5. Konkani
6. Malayalam
7. Marathi
8. Punjabi
9. Sanskrit
10. Tamil
11. Telugu

Indian Language support in the Operating Systems

Windows XP is the operating system introduced in 2001 from Microsoft's Windows family of operating systems. With built in support for over 60 scripts, hundreds of languages, and 126 locales, every language version (and each flavor) of Windows 2000 provides unprecedented support for international and multilingual computing. Windows XP expands this global support and supports multilingual computing into the client space and home computing. The users of Windows2000 / XP can do the following in their systems:

- Display Indic text;
- Input Indic text / script,
- Format and print Indic text;
- Create file names with Indic text;
- Correctly display dates, times, currencies and numbering schemes in Indic text and local cultural format;
- Correctly sort and search on Indic text;
- Set default browser language to an Indic language

Developers and programmers using Windows2000 / XP as their operating system could easily leverage this Indian Language support in their applications:

- The data format being used for a language, or the character specifications and encodings
- Date and Time formatting
- Numeric and Currency formatting
- Weights and Measurements
- Rules to sort words of a particular language

From an application developer's perspective, the system setting is like a key from where he can obtain the current cultural preferences of the user. Obtaining this information is a key step towards localizing an application.

MS Office and Indic

MS Office suite comprises of a few tools used extensively by knowledge workers and most PC users. MS Word is the word processing program of the Microsoft Office suite that helps create documents and reports. Excel develops spreadsheets that display data in various tabular and visual formats. Power Point creates multimedia presentations to display information in a graphical format. Access is a database program that stores information that can be manipulated, sorted, and filtered to meet your specific needs. FrontPage allows one to create professional-looking web pages for the Internet. Outlook is a personal information manager and communications program.

Indian Language support in these tools goes a long way by assisting the users use Indic language for:

- Automatic language detection
- Date, Time & Number Style
  - Based on Regional Settings
- Automatic Font Linking
- Editing Indic Content
  - Find & Replace for Indic Content
  - Spell Checking

- Checking correct sequence of characters
- Sorting – based on Indic Text
  - Auto Filter
  - Custom List
- Create Graphs with Indic Text
- Currency Functions
- Sending & Receiving Indic Content
- Date Formats
- Type appointments in Indic Language

Office Hindi

Microsoft launched its first offering developed specifically for the Indian market - Office Hindi. The product includes a Hindi language interface and supports nine Indian languages, empowering Indian users to leverage the global, standards based Office applications suite in the language of their choice.

The Professional version offers a switchable interface, providing users with the flexibility of easily deploying the Hindi or an English interface as required. The product also offer multiple keyboard options: Transliteration, Inscript, Remington and Godrej, along with other variations.

Indian Language Application Development using.Net

VS .NET has support for Indian languages using Unicode. One can use VB.NET, VC++.NET, C#.NET, etc. to develop any application for the .NET framework. The application developed can be deployed on any system which has the dot net framework installed. Since the environment is Unicode one can have Indian language text in Menus, textboxes, error messages, etc. One can have web-services in Indian language. There could be a web-service running from a server that provides news headlines in Indian language

SharePoint Products and Indic Support

SharePoint Products and Technologies facilitate collaboration within an organization and with partners and customers. Using the combined collaboration features of Windows SharePoint Services and SharePoint Portal Server 2003, users can easily create, manage, and build their own collaborative Web sites and make them available throughout the organization. SharePoint Services is data management and analysis software that will deliver increased scalability, availability, and security to enterprise data and analytical applications while making them easier to create, deploy, and manage. SharePoint Services can also be used as a development platform for creating collaboration and information-sharing applications. SharePoint Portal Server 2003 is a secure, scalable, enterprise portal server built upon Windows SharePoint Services that can be used to aggregate SharePoint sites, information, and applications into a single portal. All the above can be achieved in the 11 languages supported by Windows XP SP2.

Indic Support on SQL Server

SQL Server is a relational database management and analysis system for e-commerce, line-of-business, and data warehousing solutions. In SQL Server 2000, the data in Indic languages can be stored using the data types like nchar, nvarchar and ntext. These data types are defined to be equivalent to UTF-16 Unicode. Unicode data is stored using two bytes per character, which is one byte per character in the case of character data. The sorting of Unicode data is handled based on the collation name selected for the database. In the case of Hindi language the collation name can be selected as "Hindi_BIN" (BIN standards for Binary Sorting).

Technology behind Indic Support in MS products

To accommodate multilingual computing and to properly display text in a multilingual context, Microsoft introduces new technologies:

Open Type fonts: This new font format, developed jointly by Microsoft and Adobe, is in fact an extension to the TrueType font specification. The Open Type fonts in Windows 2000 include expanded repertoires of glyphs to accommodate Pan-European, Arabic, Hebrew, and Thai scripts.

Font fallback: This mechanism, made available through Uniscribe, provides a fallback font (or a default font) when dealing with complex scripts. If

the selected font face does not include any glyphs for the complex script that is about to be displayed, Uniscribe selects a default hard coded font for the given script. For example, if you have Hindi text and the font is Courier, then Uniscribe will use the Mangal font. This technique is internal to Uniscribe and developers can not add additional fonts to the list of fallback fonts.

Font linking: This technique is mainly used to accommodate East Asian languages and uses a registry setting of fonts linked in a list to a face name. If the required glyph is not in the nominal face name, GDI searches each font in the list of those linked to the face name until it finds one with the required glyph. This new font is then linked to the nominal font (base font). Users and developers should not modify the list of linked fonts in the registry, since it might cause serious display problems.

Uniscribe: Uniscribe, the layout and shaping engine for these scripts, is built into Windows 2000/XP and provides consistent support across its clients (Windows 2000, Microsoft Office 2000 and Office XP/2003, Microsoft Internet Explorer 5.0 and beyond).

Data Types used Indic Languages in various products

- SQL Server 2000 – NVarchar, NChar, NText
- MS Access – Text
- .Net – String
- C, C++ - wchar, LPWSTR