15. Summit on Internationalisation of Web
August 24-25, 2006 - Hotel Grand Ashok, Bangalore

Dr Anup K Pujari, Secretary-IT & BT, Government of Karnataka inaugurated the Summit and dwelt on how new technologies are fast changing the way Web is used. He emphasized the need of taking the Web to common man and this could be achieved by making the Web available in the local language i.e. the language of the common man.

Richard Ishida, W3C ? Internationalisation Lead introduced the importance of localization in his presentation. He highlighted the various aspects of internationalisation and localization. He introduced the components of localization and internationalization.

Swaran Lata, Addl Director, Department of IT said that aim should be to bring a visible change in the way people access the Web in all the 22 constitutionally recognized languages. She mentioned that the Indian languages are very complex in nature particularly when it comes to the development of

technologies for these languages. She also mentioned that the internationalization and globalisation go hand in hand. DIT became the member of UNICODE and W3C with the objective of proper representation of Indian languages in the Web Technology Standards and related standards.

R K Verma, Head W3C India Office elaborated on the functions of the W3C India Office and also emphasized the need for content creation in local languages and proliferation of W3C standards in India for making Web available from anywhere, anytime and anyplace for everyone.

Dr Srinivas Padmanabhuni, Principal Researcher, Infosys said that two major areas of W3C are Globalisation and Localising a Globalised application in local language. He elaborated on the concept Web Services or Service-Oriented Architecture (SOA). SOA enables ease of interoperability of various components without significant cost of interoperation/ customization. SOA is an approach to distributed computing that is loosely coupled, protocol independent, standards-based, where software is accessed as a service, where software resources interact over a network according a contract.

R Hari, Project Manager-OSSRC, IBM Software Group emphasized that the Indian languages are phonetic in nature. What you write is what you get. 5-6% Indian languages are English enabled. India ?s share in online user is 2.5%. He introduced the terms of internationalization, Localisation and Globalisation. He further relaborated that Globalisation is the process of developing, manufacturing and marketing software products that are intended for worldwide distribution. He further explained that a Globalised product is one that can handle multiple languages and cultures simultaneously and provide capabilities for customers and business partners to create global applications.

V N Shukla, Director-Special Applications, CDAC, Noida mentioned that Internationalisation is the process through which we are trying to localize something. It is a standardization process and we need to have a clear-cut roadmap for this. We should first find out what are those issues which reflect our culture. There cannot be any authority in languate which says that you need to follow this only.

Sasikumar, Sr Research Scientist, C-DAC, Mumbai mentioned that i18n is the process of designing applications so that it can be adapted to various languates and regions without engineering changes. He further elaborated that an Internationalisation software is the same executable that can run worldwide textual elements are not hardcoded. Supporting a new locale does not require recompilation of software. Culturally dependent data appear as relevant to respective region/language. The things which are affected are Text messages, GUI, Dates, Time, Numbers, Measurements, Colours, Icons, Currencies, Phone Numbers, Postal Addresses, etc.

Dr B Mallikarjuna, Academic Secretary, CIIL, Mysore mentioned that massive number of people involved in the use of multilingualism census has 10,400 raw returns, rationalized into 1576 mother tongues, further rationalized into 216 mother tongues grouped under 114 languages. Linguistically, India is made of many mini-Indias. He shared the experience of linguists in dealing with the concept of internationalization.

Dr Girish Nath Jha, Assistant Professor, JNU said that India has more than 1600 languages and 22 national languages. There are various Indian language families

◗ Indo Aryan, Dravidian, Munda

◗ Indo Aryan, Dravidian, Austro Asiatic, Tibeto Burman Addamanese

◗ Indo Arya, Dravidian, Austro Asiatic, Tibeto Burman

He further mentioned that over 95% of Indian population speaks one or other of scheduled languages. Hindi is spokedn by 40.22% people. He listed out the various issues for internationalization:

◗ Alphabet Standardisation

◗ Unicode/font issues

◗ Morphology/syntax

◗ Usage conventions

Ravindra Kumar Sr Director, CDAC, Trivandrum spoke about locales and multilingual naming. A locale provides cultural conventions according to the language and territory. CLDR is the largest and most extensive standard repository of locale data. This data is used extensively for software internationalization and localization. Multilingual Naming System is the extension of internet naming system to include the diversified alphabets and languages used by most people of the world. It includes Internationalised Domain Names (IDNs). He also spoke about visual spoofing and inscript spoofing.

Manoj Annadurai, CEO, Chennai Kabigal spoke about visual spoofing. Visual spoofing becomes possible when different domain names give rise to a similar visual representation (glyphs). By normalizing different strings and converting them into a standard representation of the visual firm, this problem could be completely eliminated.

Kewal Krishan, Technical Director, NIC spoke about the e-Governance standards particularly e-Governance localization of applications and language technology standards. He brought out the various issues in Indian Language Computing. The broad objectives of the e-Governance Committee are: Adoption of uniform encoding standards for Indian languages; Support for inputting mechanism; Multilingual support to be ensured across platforms and applications; and Websites & emails in Indian languages.

Richard Ishida explained the various practices, which should be followed while designing multilingual websites. He stressed that the standards like UNICODE and W3C standards should be followed while designing multilingual websites. He also emphasized that the needs of scripts and languages should be brought to the attention of W3C and other standards and tool developers. For making styling more interesting, CSS style sheet should be used, he said. He also mentioned that one should consider separation of content and presentation when adding scripting.

Wipro Technologies has brought out the challenges being faced for development of multilingual websites:

◗ Lack of consistent technology and process

◗ Ineffective content management

◗ Differing user behaviours and expectations

Some of the best practices brought out were:

◗ Build the website for different connection speeds

◗ Build an effective Global Gateway

　◆ Landing pages for first time users

　◆ Prominent navigational elements on each page

　◆ Located URLs

　◆ Accommodate user?s language preferences

◗ Be aware of localization issues

　◆ Colors

　◆ Numbers

　◆ Names

◗ Design navigation that is suitable for

　◆ Translated content

　◆ The number and depth of links

Richard Ishida brought the concept of International Tag Set (ITS). He said that the ITS Working Group at the W3C is currently specifying markup of this

kind that can be used by any schema developer to support international use of documents and also effective localization of documents.

Ramakrishna Reddy brought out the importance of standards while using internet. He said that the various tools for internet are ? Validators, Libraries, Parsers, Web agents, viewers, authoring tools, server side tools. He also mentioned that most of the FLOSS websites are 118N enabled.

Prof Pat hall brought out why standards are needed, what are the standards, who makes them and what benefits do they bring. He also spoke about the standards in software i.e. 118N and LION.

Rajesh Thakkar spoke on Web Services Internationalisation. He said that the Web service is a software system designed to support interoperable machine-to-machine interaction over a network. He mentioned that a web service has an interface described in a machine-processable format and other systems interact with web service using SOAP messages, typically conveyed using HTTP and XML. He also mentioned that Web services aim to be an interoperable, platform-neutral means of invoking logic over the web. He also pointed out that WS-118N is an emerging standard to provide a framework for Web Services Internationalisation.

Dr D.R. Shukla presented Land Record Information System developed and deployed by NIC across various States. He said that the system has nearly 430 million land records. He also mentioned that the data with regard to land records has been stored at the State Level in regional languages. He said, presently, NIC is in the process of converting the legacy data to UNICODE so that complete land record data of the entire nation can be converted to one uniform international standard. He also highlighted the issues relating to conversion of data from ISCII to UNICODE.

Finally, the leading panelists drawn from the Government, Academia and the Industry viz. Swaran Lata, DIT Richard Ishida, Srinivas Padmanabhuni, Girish Desai, Project Manager, Jataayu Software and Sujatha Visweswara, GM & Practice Head, Wipro Technologies put their brains together and brainstormed on the Internationalisation Strategy for the W3C. The key points that emerged out of the discussions are:

1) Constitution of an Expert Group for Development of Multilingual Websites as per W3C recommendations.

◗ Evolvement of Guidelines
◗ Development of Sample W3C compliant Website in Hindi
◗ Constituents:
  ◆ Wipro Technologies (Ms Sujatha Visweswara)
  ◆ Microsoft (Dr A Kumaran)
  ◆ TCS, Mumbai (Sh. S. Hariharan)
  ◆ CDAC-Pune (Sh. M.D. Kulkarni)
  ◆ Infosys (Dr Srinivas Padmanabhuni)

2) An expert group of computational linguists and industry to be formed
◗ Identifying the gap areas in language technology
◗ Requirements from the Government
◗ To push Internationalisation/Localisation Initiatives.
◗ Constituents:
  ◆ Industry
◗ IBM, Microsoft, TCS, Jataayu
  ◆ Computational Linguists:
◗ Dr Mallikarjun, CIIL, Mysore
◗ Prof Girish Nath Jha, JNU
◗ Mr Raymond Doctor, CDAC-Pune

3. Localisation in e-governance
◗ e-Governance, Developers Workshop
◗ 2-day
◗ Scope:
  ◆ Internationalisation
  ◆ Globalisation
  ◆ Localization
  ◆ ISCII
  ◆ W3C
  ◆ UNICODE
  ◆ How to develop an Internationalisation Application
  ◆ How to localise this Application in Hindi

Courtesy
Mr. Amit Chadha
MAIT, New Delhi