Presentations
3.1 Why Standards are Important

Mr. Pat Hall, Open University, U.K., Email: pavhall@open.ac.uk

Why standards are important

Pat Hall
Global Initiative for Local Computing
& the Open University

Examine a number of case studies to answer:
what is the value of standards?
why is this so hard to understand?
how can standards be developed?

- specification standards, not procedural standards or guidelines

---

1960s keyboard, code, fonts

- keyboard
- physical
- keys
- one-to-one correspondence
- tactile
- braille
- sounds

---

1950s connecting equipment

- tape reader
- keyboard
- card reader

- photos and scanneds
- conform to standards
- (e.g., IBM, DEC)

- electrical signals
- conform to standards
- (e.g., ASCII)

---

1970s Independent input, code, output

- input device
- output device
- rendering or output generation
- internal codes - UNICODE
- communications / applications

---

standards enable:
- interconnection of components from different suppliers
- substitution of components
- an open market

- could inhibit innovation
- but can evolve
- determined by technology of the time
No standard code in communication

Interpreted and compiled view of technology

<1990s Desk Top Publishing fabrics

Independently implemented
Require deep expertise
Continue to evolve with technology

Non-Latin character code

presentation Layers

- Public code
- Interpreted by technology
- Need extra knowledge
- Require deep expertise

standards
3.2 Usability Evaluation of Localised products

Mr. Pat Hall, Open University, U.K., Email: pavhall@open.ac.uk

Usability Evaluation of Localised Products

Pat Hall, Shailey Minocha
Open University UK

interface design guidelines

- 'universal' guidelines and heuristics
  - Nielsen, Shneiderman
- localisation issues
  - icons
  - colours
  - symbols
- extra features
  - locally-significant brand identity
  - linguistic cues

new or localised software

- common issues
  - is interface understandable?
  - does it impose cognitive burdens?
  - does the software do something useful?
  - can you do what you want to?
- common approach
  - requirements analysis
  - interface design guidelines
  - usability evaluation

functional localisation

- legal practices
  - eg payroll
- privacy policies
  - eg warehouse men knowing prices
- local work-practices
  - multiple discounts

requirements analysis

- who are your users?
  - what language, what dialect
  - actual user characteristics casual/regular
- what aspects of language will be supported
  - data storage and retrieval
  - menus and messages
  - help and user guides
  - software maintenance
- how requirements are elicited
  - focus groups, interviews, observations, surveys
  - prototypes, storyboards

evaluation

- what to evaluate?
  - prototype, partial system, full system
- select users
  - from target community
- how to evaluate
  - questionnaires
  - observations
    - think aloud, retrospective protocols, video protocols
  - usability lab
    - video cameras, audio recorder, eye-tracker
Conclusions

- eye-tracker recording
- typical usability lab
- portable usability lab
- camera
- remote user
- cloud
- data
- PC

Eye-tracker

For detailed analysis of a website
3.3 Localization Needs Language Engineering

Localisation Needs Language Engineering

Pat Hall
Global Initiative for Local Computing & the Open University

Presentations

NalRaLEC – to be funded by the EU
OpenU, LeuvenU, WarwickU, ELRA, MPI, Tilkince\textsuperscript{7} for Nepal, even though not underdeveloped

- Nepali National Corpus
  - text (3.3M words, 114K parallel) and speech (435M+194K words)
  - fulltext in XML format
  - to be released for free

- NalRaLEC dictionary – aiming at 100,000 words
  - word list and corpus for most frequent

- Linguistic tools
  - basic office software in Nepali
  - including spell checker

- Training
  - train in schools and universities
  - computational linguistics courses in university

Nepali solution

Background of Nepal

- Nepali (biter Nepal) official language since 1750
  - anglophone of Indian origin
  - 1600-1800 'cru nation, cru culture, cru language'
  - around 168 other languages
  - Nepali now almost universally spoken
  - and in neighboring regions
  - 6 other languages with written traditions
  - 36 completely unwritten

- 1990 mandated mother-tongue education
- 2003 translating mother-tongue teaching
- but how to teach the money for it?

Nepali solution

Current situation

- Nepali linguistic
  - established as native little (1670) via till.
  - dictionary at 10,000 words in 1981
  - project to 2007 run out of money

- Archive Library of Modern Purneek Pustika\textsuperscript{8} (MPP)
  - large established traditions
  - native newspapers and journals
  - young literature

- DTP since the 1980s, and have Mac and PC forks
  - 1981 movement to establish standards
  - Unicode using Devanagari, different collision sequences
  - Software localization: begun around 2005
  - support from MPI directly and via Perl for
  - Linux, compile of UTF tools, some office software

Nepali solution

Challenges

- standard localization for clips, mouse, etc.
  - how to get windows and mouse?

-导师 ‘Sanskrit’
  - not used as national language
  - subjects available from overseas
  - in local data using historical tables, takes six months
  - any attempts at versatility

- appropriate local software
Conclusions
3.4 Localisation
Prof. Jitendra Shah, CDAC Mumbai, Email: jitendras@ncst.ernet.in

Localisation: Community Act
- Present Scene in Community OS: InLinux (Red hat)
  - Small scale industry, one man or a small team
  - Government effort (Janabhaaratii): Cash infusion
  - Prolonged nascent
  - Web/CD publishing (alpha/beta) under GPL like
  - No mass usage: No records of mass feedback
- Present scene in proprietary OS: MS
  - Corporate effort: cash infusion
  - Enlisted experts: Good web site
  - Commercial Product launches

Localisation: Translation
- Ingredients: Apart from PASSION
  - Linguistics and linguistic
  - Technical
  - Community acceptance
  - Consistency
  - Testing
  - Feedback
  - Revision
  - Product
  - Branded
  - Community ownership

Dialects and Languages
- Neo-rich Vs Rich community
  - Like McDonald Vs Idli
- Access Vs diversity
- Variety Skins: Equivalence among flavors
- Factory production and cottage industry
- Locale change: on reboot or on the fly

Standardisation
- Without killing diversity: Allow assimilation
- Standard Specification: After adequate exposure
- Language Circle: Linguists/technologists/language experts/Users
- Institutionalisation:
  - Federal structure: State subjects
  - Official Language Authority:
    - Administrative usage
    - Supreme Court order
  - Welfare State Action
  - State bodies: statutory or conventional (tecb/finelessness?)

Janabhaaratii Project
Localisation of Free/OpenSource Software: Development, Deployment and Community Building

Your folder includes a pamphlet
3.5 Computerisation of Land Records
Mr. D.R. Shukla, NIC Delhi, Email: drshukla@nic.in

A Flagship E-Governance Project of Government of India

The Project

LRE
- Project Venture: NIC & MDO
- More than 657 Districts & 1 Union Territory
- Distribution of ROs in 1995 Telikova
- Expenditure approx. Rs. 250 cr.
- Began in 1988-89

Back-End
- Land Records System
- ISO 9001 Certified

Multipurpose Cadastral Data Base

<table>
<thead>
<tr>
<th>Land Act &amp; Reforms</th>
<th>Native Language</th>
<th>Available means of network access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Reform</td>
<td>AvailableMeans</td>
<td>Categories &amp; no. of ownership rights</td>
</tr>
<tr>
<td>Land use</td>
<td>AvailableMeans</td>
<td>Output formats &amp; Workflows styles</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Land Registration Categories of Land</td>
</tr>
<tr>
<td>Land Revenue</td>
<td>AvailableMeans</td>
<td>Land &amp; Property Transaction &amp; Modification</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Verification &amp; Authentication of Every Task</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Simple &amp; User-Friendly Interface</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Data Backup &amp; Security</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Data in Regional Language</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Security Using Biometric Devices</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Interactive Pointing Utility</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Online Mutation Updates</td>
</tr>
</tbody>
</table>

ACCOMODATING REGIONAL VARIATIONS

<table>
<thead>
<tr>
<th>Land Act &amp; Reforms</th>
<th>Native Language</th>
<th>Available means of network access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Reform</td>
<td>AvailableMeans</td>
<td>Categories &amp; no. of ownership rights</td>
</tr>
<tr>
<td>Land use</td>
<td>AvailableMeans</td>
<td>Output formats &amp; Workflows styles</td>
</tr>
<tr>
<td>Land Revenue</td>
<td>AvailableMeans</td>
<td>Land Registration Categories of Land</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Land &amp; Property Transaction &amp; Modification</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Verification &amp; Authentication of Every Task</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Simple &amp; User-Friendly Interface</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Data Backup &amp; Security</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Data in Regional Language</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Security Using Biometric Devices</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Interactive Pointing Utility</td>
</tr>
<tr>
<td>Land Use</td>
<td>AvailableMeans</td>
<td>Online Mutation Updates</td>
</tr>
</tbody>
</table>

Indicative display of CLR profi.
- 430 million land records
- 329 million hectares area
- 142 million hectares irrigated
- 108 million land dependent
- 63.48 millions are wastelands
- 50 million hectares are degraded (treatable)
- 18 percent of world population
- 2 percent of geographical area
PRESENTATIONS

DECODE TO UNICODE

MIXED ASCII, 7-BIT & 8-BIT

...
Needs to be Addressed

Presentations

1. How to UNICODE to ISCII
2. How to UNICODE and Back to ISCII
3. Support in Linux
4. Another

File:
   - Unicode Format Download for Web
   - Adobe PDF to handle another
   - Handle one

Translation of one language to another
Support in Linux

- Automate form download for WEB pages
- Use Linux to render HTML
- How to incrementally translate one language

Translation of one language to another

Unicode

- UTF-8 to UTF-16
- UTF-16 to UTF-8
- UTF-32 (Big Endian/Small Endian)
- UTF-16 Byte Order Mark

Selected ISO 17999:2008 and ISCI

How to encode the code point U+1D7F in ISCI
Presentations

Thank you for your time. I hope you found our presentation informative and valuable.

Thanks for your attention and valuable feedback.
3.6 Developer's Perspective
Mr. Devendra Joshi, Summit Info. Tech., Email: hrd@summitindia.com

Developing Internationalized Applications
A Developer's Perspective

MUSHASHI
- Do not develop an attachment to any one weapon or any one school of fighting
- Practice and observe reflectively
- Win

The Cooperative Game Principle
- Software development is a (resource-limited) cooperative game of invention and communication.
- The primary goal is to deliver useful, working software. The secondary goal, the residue of the game, is to set up for the next game.
- The next game may be to alter or replace the system or create a neighbouring system.

Essential and Accidental
- Essential
  - Storage (Encoding)
  - Rendering
  - Collation
  - SpellCheck
  - Thesaurus
  - Word Break
  - Transliteration
  - Localization
- Accidental
  - Choice of Tool
  - Font vs storage
  - unsupportive environment
  - Reuse?

Players
- Developers, Testers
- Coaches (Language, Design)
- Customer
- Environment
- Tools
- Standards

Breakthroughs
- Tools
- Object Oriented
- Unicode
- icuil
- Fonts
3.7 On Demand Business  
Mr. Tanveer A Faruquie, IBM India Research Lab, Email : ftanveer@in.ibm.com

IBM in India: Leaders in Localization and Open Standards
- 1998: Hosted PC DOS
- 2000: Lotus Notes translated into Hindi
- 2001: Creation and release with TEL, of design information for Indian languages
- Published in TEL, Workflow Magazine.
- May function as conversion of standards.
- 2002: IBM SDK for Java, DB2 and Websphere supports 8 Indian scripts on Linux, AS/400, Windows.
- Currently: Support 9 Indian languages in various IBM Software and Operating Systems, Unix and AIX.
- To be completed next year.
- 2003: Launched "India WSC", allowing Unicode input from legacy Windows 95/98 machines.
- Building bridges among University students on Localization technologies in IBM Campus Linux Lab.

Globalization is ...
- the capability for software and systems to work well globally and their ability to enable customers to create e-business applications that work and scale well worldwide
- enable tools and services that support the creation and maintenance of global applications that are standard, convertible on demand, highly scalable, internationalized, translator tolerating.

Globalization in practice
- Combined leading across products and platforms
- Integrate components and products across borders
- Communicate in a way that can be understood
- Support local culture and language
- Enable localization and translation
- Linguistic technologies, tools and resources
- Worldwide availability of solutions and services

Supporting Open Source in India
- Open Source Software Resource Centre in association with IT Bombay and CDAC, set up in India, to build capabilities in Indian language open source software.
- Eclipse : Open Source development tool
  - Supports 8 Indian scripts on Windows, 9 or Linux.
- IBM ICU:
  - Open Source implementation of Unicode algorithms. Converters for ISCII to Unicode may be easily built with Emacs.
  - Based on Indian language specifications in TEL, Vistwarenet Magazines.
- Use ICU APIs to implement Indian language support in your products.
- Cloudscape – Open source database

Strategy: Adopt open computing as a philosophy
- Open OS
- Open Hardware
- Open Interchange
- Open Networking
- Open Publishing
- Open Data
- Open Platforms
- Open Architecture
3.8 Localisation Applications: Speech System

Mr. Vinamra Agarwal, Prologix Software Solution Pvt. Ltd., Email: vinamra@prologixsoft.com

Localisation Applications
Speech System

Prologix Software Solutions Pvt Ltd.

Relevance to India
- Focus on egovernance and government citizen interfaces
- Low PC penetration vs relatively high telephone density
- PC penetration largely restricted to urban areas
- High PC ownership cost (> Rs. 25,000)
- Inaccessibility in far flung areas
- High illiteracy

Approximately 90% of our population who need information the most do not have access to it

Text to Speech (TTS) technology

- Automatic conversion of written text into audible speech
- Enables computers to start talking in a clear and human sounding voice
- Aims:
  - Ability to handle unlimited text
  - Naturalness of speech
- Ongoing research globally for more than 60 years
- Several existing products for English, French, German, Spanish, Portuguese, Cantonese...

The need for VociTalk

- Allows greater access to information over:
  - Telephone interfaces — auditorium size increasing day by day
  - Mobiles — national being planned for a governance, finance, information, etc.
  - Personal Computers
- Target audience:
  - People unfamiliar to read text — daily professionals, elder citizens
  - People unable to read text
    - Vision impaired — over 40 lakh
    - Illiterate — approx. 30% of our population
  - With no access to a device such as the PC and whose primary communication link is the telephone

A Digitally Divided India

Customer Benefits

- Speech — A more natural way of communicating
- ‘Localisation’
- Facilitates wider reach of product/service information
- Improves Customer Service by enabling 24/7 access
- Reduces cost of offering services:
  - Telecom operator: in a city with limited broadband uptake, 8% of 3 areas
  - Modem: reducing telephony services
- Generates new revenue streams in the form of value added services and enhanced product benefits

Whatever be the language we speak, we inherently 'think' in our native language.
Prologix Software

Presentations

Horizontal/Vertical applications

- Wall service
- Customer service
--valve control
- re-engineering of processes
- overhauling

TRAI Consultation paper, 27th October, 2004

Vouch for large-scale development of technology initiatives within member organisations.

Presentations

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- India expected to have 25 crores telephone connections
- covering approx. 75% of our population by 2007
- widespread growth in mobile telephony has been led by demand from cabbies, rikshaw pullers, plumbers and professionals in urban areas, citing it as a tool for increasing incomes.

Integration

Voice Repe - and revitalise the rural economy by creating rural micro enterprises...

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- Our Customers

- Media Lab Asia
- Microsoft
- OnMobile
- South Birmingham H.E.S.

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- Features

- Natural and pleasant sounding synthesized speech
- Multiple Male/Female voices
- Supports both Unicode and EBCI Legal
- Large existing dictionary plus capability to handle abbreviations, special pronunciations
- Ability to control pitch / tone of speech
- Rule based approach / suffix control:
  - Cijn in specific numbers as individual digits (phone numbers, etc.)
  - Letter Company (spare, places, number, Company (114, etc.).
- Digital digits
- XML tags for content sensitive synthesis (eg. <Cijn>, <digit>, etc.)
- Available on Linux, Windows XP
- Availability:
  - Vaschak 3.8 for Gamel - Supports Per Hudson performance
- Vouch for English

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- Leadership through Vision

- Identified speech technology for Indian languages as an effective rental as early as 2000
- This effort involved in learning and perfecting technology
- Version 1.0 (June 2001)
- Version 2.0 (October 2004)
- Global Benchmarking: Technology today based on cutting edge techniques
- Multiple voices: Hindi, Indian Accent of English
- Other Indian Languages: Still under development

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

Showcase Application: Railway Ticket Booking On Phone

Client: - Grameen Asia Pacific, India

Develop and deliver cutting-edge software solutions for the wireless Internet and mobile industry worldwide

- Key Features:
  - Hints that voice based railway ticket booking application
  - Tapping speech directly
  - Provides 24x7 customer assistance
  - Personalised booking with the help of a high performance "Indian" language
  - Is bi-directional - "Traveler"
  - Integrated with all major airlines services nationwide in India

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- Our Customers

- Media Lab Asia
- Microsoft
- OnMobile
- South Birmingham H.E.S.

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- Features

- Natural and pleasant sounding synthesized speech
- Multiple Male/Female voices
- Supports both Unicode and EBCI Legal
- Large existing dictionary plus capability to handle abbreviations, special pronunciations
- Ability to control pitch / tone of speech
- Rule based approach / suffix control:
  - Cijn in specific numbers as individual digits (phone numbers, etc.)
  - Letter Company (spare, places, number, Company (114, etc.).
- Digital digits
- XML tags for content sensitive synthesis (eg. <Cijn>, <digit>, etc.)
- Available on Linux, Windows XP
- Availability:
  - Vaschak 3.8 for Gamel - Supports Per Hudson performance
- Vouch for English

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

- Leadership through Vision

- Identified speech technology for Indian languages as an effective rental as early as 2000
- This effort involved in learning and perfecting technology
- Version 1.0 (June 2001)
- Version 2.0 (October 2004)
- Global Benchmarking: Technology today based on cutting edge techniques
- Multiple voices: Hindi, Indian Accent of English
- Other Indian Languages: Still under development

Prologix Software

Summit on Localisation, New Delhi, December 9th, 2004

Showcase Application: Railway Ticket Booking On Phone

Client: - Grameen Asia Pacific, India

Develop and deliver cutting-edge software solutions for the wireless Internet and mobile industry worldwide

- Key Features:
  - Hints that voice based railway ticket booking application
  - Tapping speech directly
  - Provides 24x7 customer assistance
  - Personalised booking with the help of a high performance "Indian" language
  - Is bi-directional - "Traveler"
  - Integrated with all major airlines services nationwide in India
3.9 Localization Tasks

Mr. M.N. Cooper, Modular Infotech, Pune, Email: modular@giaspn01.vsnl.net.in

Localization Tasks

- Modular Infotech has been working in localization for the last 5 years.
- Have localized Application and Operating System for many clients.
- All Modular Infotech products are localized.

Technologies Needed

- Font Technology
- Translation versus Transliteration
- Building Validated Glossaries
- Software Components - Support Languages
- Standardization - Fonts, Keyboard, Glyphs
- Decision - ISCII or Unicode

Localization

- Why?
- What all is needed.
- Technologies used.
- Work need to be done.

Fonts

- Aesthetics, Proportion, Typography.
- Readability at Lower Sizes.
- Embedded Bitmaps versus Hinting.
- Font Smoothing Technology.
- Truncation versus Condensation.
- Proportion with Latin Glyphs.
- Standardization on Glyph Shapes.

What all is needed

- Local Language Enabling (OS or App).
- Validated Glossaries of Technical Terms.
- Translation Rules - Names, Trademarks etc.
- Synonyms and Word Truncation Rules.
- Localization of Images and Sound.

Translation versus Transliteration

- Translation introduces Unfamiliar Terms
- Transliteration pollutes Language.
Thank You

OCR
• More effort should be directed to TTS and
  high-quality, readable fonts to be made
• Application
  even local language; speech enabled
• 20% to 35% population can not read/write

What needs to be done
3.10 The TILP Certified Localisation Professional Programme

Mr. Reinhard Schaler, Localisation Research Centre, Ireland Email: reinhard.schaler@ul.ie

The TILP Certified Localisation Professional Programme (CLP)

Become a certified localisation professional
Acquire the skills to develop your career
Hire professionals with confidence

Reinhard Schaler
www.tilponline.org

On second thoughts...

- Employers
  - Recruitment of staff: global yardstick, criterion for screening
  - Staff motivation: development and retention of staff

- Employees
  - Professional recognition: lack of formal and relevant qualifications
  - Mobility: not tied to national academic qualifications
  - Career: planning and development

- Training providers
  - Independent: transferable certification
  - Differentiation: quality mark
  - Relevance: academic and professional

Who is TIPL?

TILP is an institute of members of the localisation community.
TILP is owned by its individual members.

Not for profit – Global reach – Elected Council – Associated/Professional Membership

The CLP Programme

- CLP is a globally operating professional certification system, based on requirements established in consultation with key industry players
- CLP offers entry-level qualifications and a career path for professionals
- CLP certification allows professionals greater mobility and provides companies with a globally operating professional standard when hiring
- TILP offers CLP accreditation to audited courses and training providers

The Certified Localisation Professional

Initial reaction: Just another brick in the wall...

In localisation, the only constant is change.
Every project is different.
Each employer has their own job space.
There are no standard skills requirements.

We don’t need no education.
We don’t need no staff manual.
No desk under the climbers.
The client learned from the client.
Poor quality, have three files down.
All is all it’s just another brick in the wall.
All is all you’re just another brick in the wall.

Overview

- Professionals – three streams (vertical and horizontal development)
  - Engineering
    - Software Engineer
    - Quality Engineer
  - Project Management
  - Linguistic
- Two levels of certification
- Certification by accumulation of credits
  - Core modules
  - Elective modules
Presentations

Conclusion

Next Steps

LINKS

Procedures and Fees

The Institute of Collaboration

www.instituteofcollaboration.org
TEL: 352-627-3249
Fax: 627-3249

Payment

18 June 2015

The Institute of Collaboration

Procedures and Fees

1. Payment

2. Request

3. Payment

4. Request

5. Payment

6. Request

7. Payment

8. Request

9. Payment

10. Request
3.11 Aspects of Localisation

Mr. Reinhard Schaler, Localisation Research Centre, Ireland, Email: reinhard.schaler@ul.ie

Aspects of Localisation

A ‘nice to have’?
A ‘must have’?
An instrument of globalisation?
A fundamental right?

Localisation

- the linguistic and cultural adaptation of digital content to the requirements of a foreign market.
- the provision of services and technologies for the management of multilingualism across the digital, global information flow.
- [...] the commodification of translation services.

There is one localisation industry

The rationale

Three underlying principles

Motivation: Increase return on investment (ROI)
- Adapt an already developed product specifically to the requirements of a foreign marketplace with minimum effort.
- Sell it and help free resources for other purposes as the original product still is in a viable state to be sold.

Use globally acceptable content (LCD / IJON)
- Deploy products using the same interface everywhere (LCD).
- Use user-defined content, e.g., local or specific.
- Use localisation services, e.g., translation, cultural, or regional.

More localisation = Higher potential returns
- Increase the localisation effort to translate
- Reduce the resources and the information and maintenance costs of the product or service

Re-use (revenue) = As much as possible (L2ON)
- Re-use as much as possible - localised as little as possible.
- Minimise translation effort.
- Limit changes to an absolute minimum - eliminate additional effects

The localisation factory

- Process & technologies to lower localise a case study (X)
  - Current throughout: 100,000 language check-ins per month
  - 2 million lines per month
  - 50% of words leverage
  - Average time to process a file: 45 seconds
  - Fully scalable "add-a-box" model
  - Simplicity of all 30 languages
  - International version testing before US release
  - Reduced no. of release engineers (20 → 2)
  - Resulting in 75% savings per year
  - Positive ROI within 1 year

Digital content developers

Company internal stakeholders - principal aim: increase ROI

Shareholders
- Direct responsibility metrics, particularly distribution strategy

Operations
- Self-sustaining model
- Paid-for SLA, both in partners (other stakeholders and pressures)
The LRCs activities

Presentations

Highlights
Definitions

Education

 localization

Research

Projects

Focal Point
The evolution of past experiences is often mythologized, leading to a misunderstanding of reality. The myth is that...
Presentations

Witnness and Indicators
The market dictates the language usage.
When is localization successful?

Examples: Invert

Business metaphors

Invert
Thank you!
3.12 Localization Data

Mr. Mark Davis Chief SW Globalization Arch., IBM, President Unicode Consortium
E-mail: mark.davis@us.ibm.com

Localization Data

Mark Davis, PhD
Chief SW Globalization Arch., IBM
President, Unicode Consortium

Focus on SW Localization

- Unicode
  - Universal character encoding
- CLDR
  - Common Locale Data Repository

Can’t be in person, so...

Uncode: Universal Character Encoding

- Unique character codes for all languages

Importance of Standards

- Products developed in each country
  Interoperate with other products:
  inside and outside that country
- Mechanism for countries / Industries to promulgate best practices

Common Locale Data Repository

- Relatively new project: 2004
- Hosted by Unicode Consortium
  - http://www.unicode.org/cldr/
- Goals:
  - Common, required SW locale data for world languages
  - XML format for effective interchange
  - Freely available
Media Perspective of Indian Language Technologies

National Summit on Localization
Session- 14 December 10, 2004

V.S.Raja,
Group General Manager
Dainik Bhaskar

Language Newspapers Today Use...

Integrated Editorial Management Systems which cater to all aspects of content, workflow and management of editorial, advertisement and library systems at all centres and consist of centralized or de-centralized systems with seamless integration to enable users to access all authorized data through customized and pre-determined paths automatically.

The Milestones...

Technologies

- < 1978
- 1978-1985
- 1985-1990
- 1990-2000
- 2000>

- Hot Metal Era
- Proprietary Systems
- Apple Based Systems
- IBM & Apple
  (Open Systems)
- Content Management Systems

The Power of Integration

- Brings order
- Defines hierarchy
- Records performance
- Eliminates duplication
- Optimizes resources
- Reduces effort

The Milestones...

People

- < 1978
- 1978-1985
- 1985-1990
- 1990-2000
- 2000>

- Compositors
- Operators
- DTP Operators
- Journalists

The Power of Integration

- Focused approach
- Reduced errors
- Quicker pagination
- Higher security
- MIS at finger tips
- Automated workflow
The Power of Integration

- Higher Productivity
- Paperless Office
- Planned Working
- Latest Coverage
- Shorter Deadlines
- Empowered Editorial

A Sneak Preview...

Presentation for
- Overview of The System
- Reporters
- Desk Incharge
- Sub Editor
- Reports
- Pagination Facilities and Process
- About 4cplus

The Power of Integration

- Enhanced and instant communication
- Uniform & consistent quality
- Better transparency
- More control & security
- Integration with the net

How do we operate in the new system

- Publishing centres make all the pages
- Sharing the data across the centres
- All wire photographs processed at one location
- All national/international stories edited at one location
- All special stories shared online across the group.

Desk In-Charge

Story Assigner: All the Stories related to his Desk Both from Agency & Centers automatically come under this basket. (For Example: Desk In-charge of General Desk have all the story mark NAT, INT from Agency), Desk in-charge can assign the Stories to his Subordinates. In Story Assigner stories status can be identified by color.

Desk In-Charge Lookup Window

Story(s) are coming from agency(s)/bureau(s)/other printing centers by passing some specific filtration standards have information of heading, arriving time, file name and The story(s) assigned to.

- Pending Story(s).
- Assign Story(s).
- Delete Story(s).
- Pending Story(s).

- In Process of Assigning(s).
- Assigned story(s).
- For Approval story(s).
- Approved Story(s).
Assign Palate
- Team list - to whom story(s) going to assign.
- Date popup - to select date.
- Priority popup - to select the priority of the story.
- Assign Delete Pending button - to make action as per story(s) check option selected.
- Previous Data popup - to get back date unassigned story(s).
- Desk popup - to select another desk.
- Transfer button - to transfer story(s) to another desk.
- Picture button - to view pictures.

Bureau Stories
- User can search the bureau stories by heading wise, storywise, sourcewise, deskwise, filename wise etc.
- Click here
- The bureau section displays the stories received from other centers on a daily basis. The bureau interface is similar to the Agencies interface. Like the agency browser you can search in the data received from bureau centers.

Search Palate
- Source list - to which source to get story(s).
- Agency/Center list - to select option.
- Heading Containing - Search in Heading.
- Story Containing - Search in story contents.
- Author - search for particular author.
- Operator - by selecting it user can get filtered information.
- Search button - to make action against above selection(s).
- Clear button - to clear all selection(s).

In Story Assigner stories status can be identified by color.
(A) If Story has been Assigned to some one that story’s color will be Blue.
(B) If Story is Submitted for approval by assigned person that story’s color will be shown as Yellow.
(C) If the Stories have Approved by Desk in-charge that stories color will be Green.
(C) If the Stories is in process of assignment by Desk in-charge that stories color will be Orange.

Agency Stories
- User can search the wire stories by heading wise, storywise, sourcewise, deskwise, filename wise etc.
- Click here
- This feature is useful for retrieving the News from Agency. (Like PTI, UNI, VAKTA, BHASKAR). In this feature we can Search the story by Heading, By Story. We can also know the time of arrival of story from Agency. User can see as per his requirement (Like if he want only PTI Story, he can select the PTI in source combo on top).

My Assignment
- In this Basket we can get that stories which are assignment to Desk In-Charge by himself or another. Subordinate according to rights. We can also Transfer stories form this desk to other Desk.
Approved: By using this Basket Desk in-charge can see all the Approved stories related to his desk with information of who made the story and what is the filename and cols and depth of story.

Rejected: In this Basket Desk in-charge can see all the files that were rejected by him earlier.

For Approval: In this Basket Desk in-charge would find all the stories that are submitted by the subordinates for approval. This is easier for tracking file. By using this he can approve his Desk file, which comes For Approval.

Story Submission:
- Priority: Low/medium/high
- Slug: Slug going with story.
- Category: All category comes under it.
- Sub Category: All Subcategory under selected category.
- Local Save: For local saving the story in user machine.
- Temp save: User can work on the story after some time but doesn’t want to send story for approval then he can use temp save otherwise for final save click on save.
- Page Details: Desk In-Charge can specify the particular pages in which story will flow.
- Edition: Desk In-Charge can specify the edition in which story will flow. Story will then flow into all pages of that edition.
- Group: Story can also be assigned group wise like sports, commerce etc.

Index:
- Presentation for
  - Desk In-Charge
  - Sub Editor
  - Reports
  - General showing all features
  - Pagination Facilities and Process
  - Writer to be used by Reporters and Others
  - About 4cplus

Editing Process:
Sub Editor will make any story that story will go in Desk In-charge’s “For Approval” basket for approval. Sub Editor cannot make story directly for Pagination. Sub Editor’s story can be rejected or approved by his Desk In-charge. Sub Editor can Search Center & Wire Story if rights is given to him.
My Assignment

This is same as Desk In Charge's My Assignment. Here also we can see that story which is assigned by some one to Sub Ordinate. User clicks on the story and the story opens into the editor automatically.

For Reporting

Approved Story:
This Report is basically to show approved stories in a particular desk. Desk In-charge can see only his desk's approved stories. Management People can see all desk stories, which they want. In this report we can see last date Approved.

Submitted

In this Basket we can see only that story, which has submitted to Desk In Charges for approval.

Note: Remaining all features will work as Desk In Charge’s Features.

Approved Story Detail

Clicking on the Story we can get the whole detail of story.

Employee Work

Employee Work: This Report shows all Employee Work. In this report we can see desk wise Employee Report like how many Stories has Rejected, Approved, Being Modified or New. In this report Desk In-charge can see all his Desk Sub Editor's Status Status Status.

Index

Presentation for
- Desk Incharge
- Sub Editor
- Reports
- General showing all Features
- Pagination Facilities and Process
- Writer to be used by Reporters and Others
- About 4cplus

Employee Work

Employee Work: This Report shows all Employee Work. In this report we can see desk wise Employee Report like how many Stories has Rejected, Approved, Being Modified or New. In this report Desk In-charge can see all his Desk Sub Editor's Status Status Status.
Work Summary

This Report is useful for seeing all Desk Stories Report. This Report can show all stories total in a particular desk, Total Approved Stories, Rejected Stories, Center Stories, Wire Stories. By Default this show today’s Status but we can see all wire status (Total) also in this Report. By using this Report we can Track the Status by Desk as well as by Employee also.

Center Story Arrival

Center Story Arrival: This Report shows the actual time of a story means what time story has come from Center.

Edition Status: This Report shows All Page Status Report means which page should release in what time & till now how much page has completed.

Desk Status: This Report shows the Graphical status of varies desk where this shows how many stories have been approved, new, rejected etc.

Index

Presentation for
- Desk(Charge
- Sub Editor
- Reports
- General showing all Features
- Pagination Facilities and Process
- Writer to be used by Reporters and Others
- About 4cplus

Bureau Tracker

- All stories received from bureau is being listed separately for each bureau.
- All pages received from the bureau is being listed separately for each bureau.

Wire Tracker

- News is being directly received on computer. The stories captured are converted to useable formats automatically by the system.
- Facility to define Multiple Filters.
- Supporting of both 5 bit and 8 bit format.
- Automatic saving of news received from different sources on predefined location.
The desk head is able to retrieve stories from each bureau with ease and at the click of a button. All pages sent by the Bureau module are compressed and decompressed automatically and not transmitted until all elements defined in the page are collected before output. There is two-way communication between the centers and the bureaus. All pictures transmitted by the bureaus are forwarded to the concerned desk but should not be used in the edition unless corrected by the center. User can search the bureau stories by heading, storywise, sourcewise, deskwise, datewise, filename wise etc. Click here

• Supports all formats of images (*.tif, *.jpg, *.eps, *.bmp, *.pdf, *.gif, *.jpg, *.tga) however all photographs used in the system should be in one format across the centers for output and stories.

• Pictures in the library are in the actual scanned size without cropping. Any picture used on a particular date should have information about the date of use.

• Every processed photograph has a thumbnail, a low-resolution image of actual size and high-resolution image of the photo. The system should use low-resolution images for all purposes other than output and only use high-resolution images for output.

• Option to send pictures directly to Photoshop or any other image editing software.

• Option to show the gallery caption to editor while composing own caption simultaneously.
• The user can select the desired keyboard. (Remington/Phonetic etc.)

• The text of the story is fully editable as desired.

• The number of columns for the story is being definable from a drop down menu, which is customizable by the administrator to any desired width. The preview is showing the story in the specified width.

• The point size of the Crossers is alterable as per the requirement or customization, which is displayed with proper line endings.

• The position of the picture/crosser can be dynamically changed. When the number of columns is changed, the crosser/picture adjusts to the new settings automatically.

• The system follows the style sheet of the page / edition as customized while being edited.

Continue….

• Option to switch between story writer and preview.

• Spell check facility.

• Thesaurus for spelling to take care of content for other editions where the spellings may be slightly different.

• Add word to custom dictionary option.

• Auto conversion of the story writer font into the pagination font.

• Automatic conversion of common file formats into the editorial system format.

• Ability to change the font, font size and type within pre-defined templates. The selected options are displayed.

• It displays word count, character count, line count and depth measure of the story on a page in centimeters or millimeters.

• Suggestions for the layout and style sheet are available from approved templates.
Presentations

Periodic auto save facility available on the system.

Each user / category / group is assigned rights by the administrator. The user rights can be modified by the supervisors (Desk In charges, Editors etc.). Security is through user and rights management. Each users rights can be defined / modified / deleted as per requirement by the supervisors / administrators / editors only for persons working under them.

Desk In-Charge
Sub Editor
Bureau(s)
Agency(s)
Story
Input
Assigning story
Submit story
For approval
Reassign
Approved
Story(s)
Paginator
Archive
Page(s)
Story(s)
News Paper
Page Pass
Rejected
Story(s)
Work Flow

Option to allow users to work on different articles on the same page. Stories / elements can be forwarded to the desired users / desks / categories basket / folders as required from time to time. This should be possible by the administrator or the editor. It can possible to send all stories/elements to desired destinations like users / pages / folders / centers / desks etc across the various centers.

Right Click and then open the story in new window

A matrix is provided for assigning the rights to the users by the administrator.

Permission can be assigned to individuals and groups.

The workflow is through an intra office email and instant messaging.
The system maintains the original version of the story. All revisions made are stored as versions with ascending numbers. Any version can be usable on the page. However, by default the latest version only is being forwarded to the next operation.

- The system does not allow identical headlines / story matter to be repeated in an edition and message will come when headline and slug will match while saving the story.
- Other centers can also be integrated with the editorial workflow system to allow them remote access.
- Finished page is automatically sent to the finished folder of output on image setter or center.
- Module to define the priority of the pages to be transmitted.
- Priority of pages / stories / elements for transmission can be defined as per requirement on a daily / permanent basis.

• Editions / Pages / Page no. are customized as per requirements.

To Add New Edition
To Add User the rights for particular Page of particular edition

• Complete tracking of a story, its author, its current user, the desk to which it belongs, its due date and the page to which it belongs.

Administration
• Provision to maintain the system data (fonts, logos, user names, user profiles, etc.) from the head office/important hubs.

New Employee
New Employee Desk
And Designation

• The database and server replicate in the background automatically. This does away with the need to manually transmit articles and images.

• Administrator can access any previous/other version of a story or page.

• Edit Log
• Edit Log
• Edit Log
• Edit Log

Presentations
Presentations

- Provision for stories that come for approval to be reassigned to another person.

Click on submitted story
Select Sub Editor
Click on reassign
Final click to reassign

- Maintains a complete list of incomplete, pending, overdue or deferred stories and can see any time in a single list.

New window with all stories which are pending, overdue will appear and user can select the ones that need to be changed.

- Event
  - Work schedule for the editorial staff for job tasks, leave, travel etc.
  - Monitoring of assignments and deadlines for the work assigned
  - Events that had been scheduled in the past can be reassigned by just changing the date / time

- MIS
  - The System generates comprehensive log report as required from time to time and the system is flexible enough to cater to the requirements of editorial and management.
  - The system generates the daily log report of users who logged into the system and out of it on a daily basis for a particular center or any/all centers.
  - A log of the users, indicating the time they first logged in, last logged out and time spent on the system by each user for the any/all centers.
  - Monthly Report of each user of the time spent by him on the system.
  - Monthly/Annual Report of stories generated by a user on the system.
  - Monthly/Annual Report of stories subbed by a user on the System.
  - Authorized users can view complete log of total work done for the day or period in different formats. Variables include, Total no. of stories, stories carried in various categories, page wise stories etc.
  - Log Report of rights given to a user / modified by authorized personnel on a daily / weekly / Monthly / Annual basis.

- Index
  - Presentation for
    - Desk Incharge
    - Sub Editor
    - Reports
    - General showing all Features
    - Pagination Facilities and Process
    - Writer to be used by Reporters and Others
    - About 4cplus

- Log in screen for QuarkExpress. After that user gets the right to view the page. According to rights, user will be able to see his pages of edition.
Opens the approved story for particular page of that edition. This pagination palette Floats the story on Quark page automatically as per the specification of story. Sign shows that this story is floated on the page of same edition.

Open the page of that specified page of particular edition. User selects the story from pagination palette and drop it over Quark page. The story floats and cross comes ahead of that story in palette.

Quark Page Change column utility can float the story of selected box into specified column(s) selected by user. This feature creates line according to the desired position of the box. The feature resizes the box according to size of story.
Quark Page

It resizes the box according to size of image.

Open the page of that specified page of particular edition

User selects the story having the crosser from paginator palette and drop it over quark page. The story floats and cross comes ahead of that story in palette.

Quark Xpress

• Page can be designed either before or during the writing and editing process.
• Menu driven option to move stories in a Quark Xpress page.
• Auto fitment of the story (to top left) on clicking in the page template.
• Auto flow of the crosser in the format defined in the story writer.
• Fit box to text
• Fit box to image.
• Version control of pages.
• Stories that have already been flown in to the page, are locked automatically and not be allowed to reused in the edition.
• On increasing and decreasing the columns the story is being auto adjust the depth and also auto scale the headline to fit in the new width. It also auto align the other components like intro boxes and pictures.

Continues...

• The pages designed on the system are compatible on Quark Xpress versions 3.32 / 4.0 / 4.1 as required.
• Articles / stories that are continued on another page are linked.
• Any changes made at the last minute on the quark page are automatically saved in the database.
• Having the provision for linking of stories continued on another page is automatically with proper set-offs.
• In case of continuation of a story to continuation page, any alteration / deletion made in the original page should bring the continuation story back in original page and the continuation page maker is being warned about the change.
• A page can be viewed / edited by any number of authorized users provided the page is from the database.
Clicking on the Button Opens a new document, where the user can type a new story. A new document can also be opened by Pressing Ctrl + O. The user need not go to the mouse. A new document can also be opened from the File menu.

Clicking on this button opens all the files in the Multiple File Opening List box (Right Most List box). Adds Files from the File list Box to the Multiple File Opening List box so that multiple files can be opened in one go. This new feature reduces the amount of time taken to access the network. In the Summit software the user needs to travel across the network to access say the haryana desk. Dbl – Clicking on File List Box would open the individual file i.e. a single story.

Further if the document is new then the Save As form pops up asking the user to type in the name of the new document. On Clicking this button we can save the current document in the path specified in the caption of the document. A 'SAVED' message shows up as a visual confirmation that the file has been saved. A file can be saved from the keyboard by pressing Ctrl + S button and also from the file menu.

On clicking this button we can see the print preview of the composed story. Various views of the Print page can be seen.
This is a toggle button. Depressing it hides the headline. Clicking on the button again pops up the headline and makes it visible.

Pressing Ctrl + H from the keyboard also hides all headings.

Earlier, two different softwares needed to be installed in the machines, one the Summit Editor for typing and editing the stories and the other, Summit Dongle which provided the keyboard facility. Hence the users needed to first load the SM software and then open the Editor for typing. If they needed to change the keyboard then they had to bring focus to the SM window and the change the keyboard there. In the new software all the keyboards have been embedded in the Editor itself. Hence the users need not switch between windows to change the keyboard.

Earlier (as in Summit interface), to randomly change the keyboard from Hindi to Roman users needed to key in the following combination: Ctrl + Shift + Alt + Space. Now they have to just press Ctrl + Shift + Alt. To change back to Hindi Keyboard they had to press: Ctrl + Shift + Alt + H. Now they would only have to press: Ctrl + Alt. The users here welcomed the above change.

This option box shows the different Font sizes to create a Heading. Further a Crosser can also be created through this Combo Box. Pressing Ctrl + Shift + Right Arrow Key increases the Heading Size. Pressing Ctrl + Shift + Left Arrow Key reduces the Heading Size. Hence the user again need not access the mouse to make a heading change.
This is another new feature. We have given a provision for a Comments Box. The users can type in any information about the file in this box and the contents won’t flow into Quark as well as in print.

A Comment Box can be directly created from the keyboard, by pressing Ctrl + K. Hence the user need not use the mouse.

Cols. Indicate the number of col. the story would finally flow. Start & End indicate the start & end Col. Of the headline. Again we have given a shortcut from the document to access these text boxes. Pressing Ctrl + Alt + C takes the user to the Cols text box. Pressing Ctrl + Shift + < Key reduces the Condensation Point. Pressing Ctrl + Shift + > Key increases Condensation Point.

This sets the alignment of the heading such as Left, Right or Center.

With this we can set headline or story as italics, bold or underline.

This interface allows the user to convert a RTF file into a STY File. Hence a story edited in WordPad or Microsoft Word can be converted into a STY file with this converter, and carried forward for pagination.

The Open File Interface which we have given (refer to slide no 4) restricts the user to few folders, which hold meaning to him/her. It does not allow the user to open his files from any other domain. For example a user in Haryana desk cannot open any story pertaining to any other desk. However, a Desk Incharge might need to pick up a file from some other directory. Hence this feature gives the flexibility to the Desk incharge to open files from different folders.
Interface for opening files in Diff. Folders.

Interface for saving files in Diff. Folders.

We have given a Timer utility in this application which takes Back Up of all open files every 6 seconds. The Back Ups are taken in a folder created date wise in a BackUp Folder. All Untitled's are saved as Untitled with the No following it, like Untitled0 and so on. Hence if the computer closes down accidentally, all open files would have a latest back up. To Open all backup files we click the above menu option.

While observing the workflow, we realized the users many a times, needed to open a file which they had earlier closed. This led to a lot of wastage of time, traveling through the network and looking for a particular file and then opening it. Hence we gave this feature – A Recent Files Window.

This Utility shows the total no. of files currently open.

Right click on headline title shows the alignment & style options.
INTRODUCTION

In an emerging information and knowledge driven society, 4Cplus aspires to help businesses to use “Information Technology” to enhance productivity, transparency, customer satisfaction, and humanity across the global economy.

4Cplus has grown with its customers on an emotional chord to make a lifetime partnership with them.

4Cplus provides technology solutions and services processing information and knowledge in a corporate environment:

• Enterprise Solutions (ERP)
• Media
• Sugar & Food Processing
• Distilleries and Beverages
• Manufacturing
• Content & Workflow Management
• Collaborative Computing and Knowledge Management
• Multilingual Technologies
• High End Internet Applications
• Internet Data Center
• End to end technology solutions

THANK YOU for Your Participation

The Editorial Team

NewsWrap™ - The News Lifeline

NewsWrap delivers editing, pagination, and content and document management functions by combining industry-standard desktop publishing applications with purpose-built editorial and pagination software — all connected to a powerful, standard, relational database to track constantly changing content. NewsWrap understands that any newsroom technology should take into account the entire news gathering process — from start to finish — and that it should be affordable, reliable and customizable. NewsWrap stands alone in offering tools that handle newsroom tasks outside the scope of a traditional editorial system. To that end, NewsWrap includes modules to track sources, photos, graphics and stories — as well as robust software to write and edit stories and to handle wire feeds.

NewsWrap Endorsements:

Dainik Bhaskar – India’s largest Hindi daily

Dainik Jagran – India’s second largest Hindi Daily with 19 editions, circulation of over 1.5 million copies everyday (ABC) and readership base of over 10 million readers (IRS 2001 Round 2).
3.14 Nurturing Living Languages

Mr. Mahesh D Kulkarni, C-DAC Gist, Pune, E-mail: mdk@cdacindia.com

Mobile Computing devices
- Anytime, anywhere information access.
- "Quality of Service" offerings is key success for the mobile service providers.
- Though voice communication has been leading application, non-voice services offer enormous opportunity, if it is provided in a operationally efficient way & has mass appeal.
- Non-voice applications include instant messaging, games, news, entertainment, productivity enhancement tools, customize user interfaces and handset-dependent updates.
- New voice-based applications – voice recognition and push-to-talk capabilities – are emerging as well.

Language Mission of C-DAC

'Dissolving language barriers by establishing standards; research, development & deployment of technologies and solutions; & disseminating knowledge for effective use of IT by masses’

Localization Applications: A Case Study
- Mobile handsets.
- Broadcast equipments.
- Printers.
- Pocket Translator
- POS – Indian Language enabled Point of Sales Terminal.
- Multilingual Display systems.

Server based solution for Indian language support
- Mobile phone having normal graphic LCD screen can be used.
- No need of language component on mobile handset, since all processing is done at server end.

Methodology
- Sender types Indian language SMS in phonetic English preceded by language tag (like HN for Hindi).
- Server reads the language tag & appropriately converts ASCII data into corresponding ISCII / UNICODE data.
- Conversion & composing SMS from ISCII to bitmap string done at server end
- The bitmap string is sent to the sender for verification.
- Sender verifies & confirms.
- Message is delivered to the receiver mobile handset as a picture message.
Presentations

Some more are in pipeline.

CDAC GST has provided solution to most of the

The solution...

Embedding the Indian language engine in

the mobile handset...

How to access the data?

- How to display the data?
- How to input the data?
- How to store the data?

To enable any Indian system/surveys for Indian

Indian language adaptation for BlackBerry

Indian language mobile

A requirement of Indian language mobile

A requirement of Indian language mobile
3.15 Machine Translation on Demand Tool

Mr. Tanveer A Faruquie
IBM India Research Lab, New Delhi, E-Mail: ftanveer@in.ibm.com

Machine Translation: On Demand Tool

Tanveer A Faruquie
IBM India Research Lab
New Delhi

Table: Scalable SMT Framework for Indian Languages

<table>
<thead>
<tr>
<th>Work at irl.ibm.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source-Channel Paradigm</td>
</tr>
<tr>
<td>Prototype Hindi-English bidirectional SMT system:</td>
</tr>
<tr>
<td>200,000 sentence pairs of parallel corpus</td>
</tr>
<tr>
<td>80 Million words monolingual Hindi corpus</td>
</tr>
<tr>
<td>400 Million words monolingual English corpus</td>
</tr>
<tr>
<td>Evaluation: English-Hindi</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training corpus size</th>
<th>150,000 sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of test sentences</td>
<td>1982</td>
</tr>
<tr>
<td>BLEU Score</td>
<td>0.3261</td>
</tr>
<tr>
<td>NMT Score</td>
<td>0.8296</td>
</tr>
</tbody>
</table>

Machine Translation @ ibm.com

IBM WebSphere Translation Server
- 14 language pairs
- On-the-fly translation of both static and dynamic web-pages
- Translation using a servlet or JSP
- Windows NT, AIX, Solaris and Linux
- Up to 500 words per second
- Machine translation services for multilingual email and chat

Globalization imperatives

- More language pairs will be supported
- Standardized easy to use APIs
- Lotus Translation Component: Single Java Interface to Machine Translation Engines
- Linguistic technology should converge on common architecture
- UIMA Architecture: MT engines to converge to this architecture
- Application development products should provide easy way to access MT Engine API
- Open standardized interface to allow plugging of tools from different sources

Machine Translation @research.ibm.com

- Statistical Machine Translation
  - Link #8: Source-Channel Paradigm
  - IBM Translation Webs 1-1
  - Advantages
    - Data Driven
    - Taskable
    - Easy to build a new MT system
    - Bidirectional
    - No deep skills in Linguistics required
  - English-French, Arabic-English, Chinese-English, English-Hindi

- Logic-based Machine Translation
  - Deep level of language analysis
  - Good translation quality

Machine Translation: India

- Problem #1:
  - Too many language pairs
  - A Language Barrier will continue to be a problem.
- Problem #2:
  - Fragmentation of effort
  - No standardized effort at solving MT problems

Solution:
- Problem #1: Statistical Machine Translation
- Problem #2: Collaborative work (2-3 teams)
- Problem #3: Common Tools Framework plus Standards
3.16 Indian Languages Standardization for e-Governance

Mr. D.R. Shukla, NIC, New Delhi, E-mail: drshukla@nic.in

Characteristics Needs to be Addressed

- Indian Language Standardization Group for e-Governance

What is Unicode?

Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language.

UNICODE & ISSUES

- Any character:
  - There is no separate code for Amharic script.
  - Characters of Amharic need to be added to Unicode code block if it is used for Amharic as well.
- Geho
  - The Amharic script needs to be added to each Jide script code set.
  - Currency signs in Geho also used as a symbol in the text.
- Dhivehi
  - Only 216 symbols available in Unicode.

Compatibility

- 16-bit international character encoding
- Windows 2000 uses Unicode version 2.0
- UCS-2 (null)
- UCS-4
- ASCII
Presentations
3.17 Localisation Issues in Multilingual Voter’s List
Mr. Rajesh Aggarwal, Election Commission of India, E-mail: rajeshaggarwal@eci.gov.in

- 700 million voters
- 800000 polling stations
- 4120 + 5 Assembly seats
- 15 languages
- 445 million voters issued EPIC
- # PS, electors, EPIC
- Eroll Languages

- roll-bihar-hindi
- roll-gujarati
- roll-manipuri
- URDU
Unequivalent pronunciation & spelling

- Open GIS, all 1 lac places
- Open Source phonetic matching
- Unicode-Latn, in public domain
- (place A, people) in English-ISCl-2
- Creating National Dictionary of names

**Come, join us for...**

**Problems**

- Tables
- Consolidating 36 lac places to 10000
duplication
- Phonetic matching for search and de-
- ISFCC-ISCII-UNICODE-URI
- Getting data to standardized tables
- Developments from delimited structures

**Kerala Map**

**Districts**
3.18 Terminology and Transliteration in Localization

Ms. Rekha Govil, Banasthali Vidyapith, Banasthali, E-mail: grekha@banasthali.ac.in

Terminology & Transliteration in Localization

Rekha Govil,
Banasthali Vidyapith

P.O. Banasthali, Vidyapith
Tel: +91-1492-284725/284736/284760
Fax: +91-1492-284725
email: grekha@banasthali.ac.in
URL: www.banasthali.org, www.17h4.org

Localisation

The process of modifying a system or product to be used in the local environment, language and culture of a particular region or locality is termed as Localisation.

Steps to be followed in Software Localization

- Language enabling - Emoting, rendering, input, cursor and edit key controls.
- User Interface - Menus, Dialog boxes, messages, documentation, icons, layout, Colour Scheme, Interaction models.
- Operational - Interoperability issues, documentation standards practices.
- Cultural - Behavioural models, rules.

Banasthali Vidyapith - more than two decades of participation in Localization activities

- The Department of Computer Science was established with financial assistance from DOE to run PGDCA course in Hindi way back in 1983.
- As a consequence, a series of Text books in Hindi for PGDCA course was developed in first of its kind effort.

Enabling Localization in the Software eventually leads to Software Globalization

The combined process of Internationalization and Localization of a system or product to address the global market (use).

In fact the concept of Localization leads to the process of Globalization which is defined as:

Globalization = Internationalization + NOT Localization

Banasthali Vidyapith - more than two decades of participation in Localization activities

- The Department of Computer Science was established with financial assistance from DOE to run PGDCA course in Hindi way back in 1983.
- As a consequence, a series of Text books in Hindi for PGDCA course was developed in first of its kind effort.
Presentations

Development of a Curriculum for

2. The development of the Curriculum for

3. Presentations of the Curriculum for

4. The development of the Curriculum for

CALT - Centre for the Advancement of Language

Hilfe für Kinder

Aktives, interaktives, kreative Lernen in der Kindergartenstufe

CALS - Centre for Language Learning and Teaching

Tanzkindegarten (Central Area)
A generalized framework leading to a CAE tool for localization integrating all tools under one umbrella.

In fact a place same which a right place need to be kept can be designed for Software localization. In this case tool localization and translation need to be provided for localization.

1. Interface localization
2. Application specific configuration generator
3. Application specific configuration generator
4. Application specific configuration generator
5. Application specific configuration generator

Thank you!

Rahul Goyal
raju12345@pm.in
3.19 Linguistic Resources and Access Tools for Localization

By
V N Shukla
Director (Sp App), CDAC Noida

Linguistic Resources and Access Tools for Localization

Summit on Localization
December 08 2004

Some Tools needed for Localization

- Machine Translation
- Translation
- Multilingual Word Processors
- Web authoring tools
- Spell Checkers
- Grammar Checkers
- Concordancers
- etc...

Development of these tools require different linguistic resources

Localization

Localization is the adapting of a product (or a service) to a new linguistic and cultural region. Along with the translation process, localization includes adapting to country-specific or cultural conditions.

Gyan Nidhi: Parallel Corpus

"Gyan Nidhi" which stands for "Knowledge Resource" is parallel to 11 Indian languages (Hindi, Punjabi, Marathi, Bengali, Oriya, Gujarati, Telegu, Tamil, Kannada, Malayalam, Assamese). It aims to digitize 1 million pages altogether containing at least 50,000 pages in each Indian language and English.

Sources for Parallel Corpus

- National Book Trust India
- Sahitya Akademi
- Foreign Publishing Houses
- Publications Division
- SASTI, FIFAS
- Post-Pub India

Gyan Nidhi: Multi-Lingual Aligned Parallel Corpus

Gyan Nidhi corpus consists of text in English and 11 Indian languages (Hindi, Punjabi, Marathi, Bengali, Oriya, Gujarati, Telegu, Tamil, Kannada, Malayalam, Assamese). It aims to digitize 1 million pages altogether containing at least 50,000 pages in each Indian language and English.
Thank you!
### 3.20 Standardization Strategies (Glyphs, encoding, locale, sorting, Transliteration)

Mr. Mahesh D Kulkarni, CDAC, Gist, Pune, E-mail: mdk@cdacindia.com

#### Background...
- India is a multilingual country with 22 officially recognized languages with various dialects.
- The 22 languages are Hindi, Marathi, Gujarati, Punjabi, Bengali, Assamese, Manipuri, Nepali, Oriya, Telugu, Tamil, Malayalam, Kannada, Konkani, Sanskrit, Urdu, Kashmiri, Sindhi, Bodo, Maithili, Santhali, Dogri.
- Some of them have Persian-Arabic origin, while most of them have their orthography derived from ancient Brahmi script.
- For preservation of cultural heritage, traditions, customs, languages need to be protected & well represented.
- Disappearance of the language means loss of community’s traditional culture, heritage.

#### Standardization Strategies (Glyphs, Encoding, Locale, Sorting, Transliteration)

- Indian languages are phonetic in nature & the scripts are further classified as Northern & Southern:
  - Northern scripts: DV, AS, BN, GJ, PN, OR
  - Southern scripts: TM, TL, KN, Ml
  - Perso-Arabic scripts: Kashmiri, Urdu, Sindhi, Persian, Arabic, Pashto.

#### Points of presentation
- Background & relevance of Languages for IT penetration.
- Complexities in Indian languages.
- Standards available in Indian languages
- Standards requirements
- Standards - future requirements.
- Relevance of standards.

#### Indian Scripts Origin

Background...
Indian Language Complexers

 devant les cartes avec le texte en langues indiennes.

 Semantics: a sequence of characters.

 16-bit code for Roman with

 Universal access: W3C, ISO/IEC

 a sequence of characters.

 8-bit code for Roman with

 Universal access: W3C, ISO/IEC

 Universal access: W3C, ISO/IEC

 Jammu Kangra complexers in alphabetical order.

 Indian Language Complexers

 devant les cartes avec le texte en langues indiennes.

 Semantics: a sequence of characters.

 16-bit code for Roman with

 Universal access: W3C, ISO/IEC

 a sequence of characters.

 8-bit code for Roman with

 Universal access: W3C, ISO/IEC

 Universal access: W3C, ISO/IEC

 Jammu Kangra complexers in alphabetical order.

 Indian Language Complexers

 devant les cartes avec le texte en langues indiennes.

 Semantics: a sequence of characters.

 16-bit code for Roman with

 Universal access: W3C, ISO/IEC

 a sequence of characters.

 8-bit code for Roman with

 Universal access: W3C, ISO/IEC

 Universal access: W3C, ISO/IEC

 Jammu Kangra complexers in alphabetical order.

 Indian Language Complexers

 devant les cartes avec le texte en langues indiennes.

 Semantics: a sequence of characters.

 16-bit code for Roman with

 Universal access: W3C, ISO/IEC

 a sequence of characters.

 8-bit code for Roman with

 Universal access: W3C, ISO/IEC

 Universal access: W3C, ISO/IEC

 Jammu Kangra complexers in alphabetical order.
### 3.21 Encoding is the Base for Computing

Mr. Nagarjuna G., Free Software Foundation of India, E-mail: nagarjun@gnu.org.in

**Encoding is the base for computing**

- Nothing can be done without encoding
- All data manipulation depends on this single factor
- All our time is spent mostly in either creating or retrieving encoded content
- Why am I talking about basics, don’t we all know them
- So far indico-computing community/industry/Govt did not behave as if we understood these basics

**What should we do?**

- Font encoding should be declared illegal, and in cases where it is required for technical reasons, users must be warned and informed that they are about to encode in a private standard.
- Manufacturing ICT solutions, without meeting these conditions should not be given licenses to do business in the country. Or Govt should ensure this in some other softer means if possible.

**Let us accept our failures**

1. In arriving at a public/open standard
2. Providing a computer that can speak the languages of the country
3. Possibly we are the only special country (IT super power) where the industry and government encourages private encodings
4. Providing content in our own languages

**consortium**

- A consortium of all stakeholders, similar to w3c should be formed as immediately as possible to ensure the implementation of the above points
- We should withstand the global pressure to assert our independence in matters related to patenting
3.22 Speech Technologies: Enhancing the User Environment
Mr. Nitendra Rajput, IBM India, Research Lab, E-mail: rnitendra@in.ibm.com

Speech Technologies: Enhancing the User Environment

User Interfaces with Speech Enablement

...In the Indian Context

- Dictation applications
  - Reduce the need to know typing
  - Directions adjusting to different keyboard layouts...to an extent
  - My dad is...more penetration of IT remote to services
  - Voice-based keyboarding
  - Talking ATMs
  - Telephony IVR systems for information retrieval
Thank You