5. Report on Interaction with Localization Research Centre (LRC) at University of Limerick, Ireland

**Date:** 12th November to 16th November 2007  
**Venue:** University of Limerick, Ireland

**Summary of discussions:**
- Localization Research Centre of University of Limerick, Ireland has spearheaded the localization activities by imparting training, tools showcase, consultancy, standards development & involvement of industry.
- Because of similar challenges faced in language technologies within India it is strongly felt that a similar setup needs to be established in India.
- It is proposed that the activities to be divided into two phases. In the first phase the activities need to be undertaken in project mode with involvement from LRC, University of Limerick & in the second phase, a separate centre to be formed to undertake larger activities inclusive of localization, linguistic resources, W3C, CLDR and others.

**Discussions with Prof. Reinhard Schaler**

The Localization Research Centre (LRC) is the information, educational, and research centre for the localization community. Established in 1995 at University College Dublin under the Irish Government and European Union funded Technologies Centres programme.

In early 1999, the Executive Board of the University of Limerick (UL) approved the establishment of the Localization Research Centre (LRC). This centre is the result of a merger of the Centre for Language Engineering at the University of Limerick (UL) and the Localization Resources Centre, formally based at University College Dublin (UCD).

The LRC provides a wide range of support activities for the localization industry and the professionals working in it. These activities include:

- Research - Postgraduate research support
- Localization Technology Laboratory and Showcase (LOTS) - Showcase of Tools, distribution, Infrastructure, network and support for localization community, tools reviews.
- Training
- Publication - Localization Focus - the International Journal of Localization (since 1996), Localization Directory (Yellow Pages), The LRC Reader (a collection of interesting articles from past issues of Localization Focus magazine)
- Conferences - Annual Localization Conference (since 1996),
- Awards Academic Localization Awards,

**Some of the highlights**
- Started 1996 “localization focus” newsletter, 1997 started graduate diploma in localization.
- Localization Research Centre gave services of localization free of charge for initial 3 years.
- Localization Research Centre participates in “Organization for the Advancement of Structured Information Standards (OASIS)” and to certain extent in “Localisation Industry Standards Association (LISA)” standards.
- University of Limerick has industry advisory board for Localization Research Centre activities.

**Revenue model of LRC**
- National Software Directorate, Irish Government seed money for LRC for three years.
- EU collaborative projects with industry participation (done 6-8 projects)
- Funding from Irish Government for Localization related Projects.
- Giving Consultancy to the MNCs
- Through Post Graduate programmes and PhD programs & other Training Programmes.

**EU funded project being executed / executed by Localization Research Centre, University of Limerick**

- **IGNITE** - Linguistic Infrastructure for Localization: Language Data, Tools and Standards - A project that will establish a central and accessible repository of European linguistic resources, strongly connected to and supported by the localization industry.
- **ELECT** - European Localization Exchange Centre - Providing reliable information on best practice, facilitating easy access to know-how and technology, making available guidelines on linguistic and cultural customization, and enhancing the visibility and recognition of this industry in Europe and worldwide.

**Certified Localization Professional** - Establishing an accreditation system for the Software Localization Industry.

**WEB-IT/EFCOT** - A web-based terminology database for Information Technology/European Forum for Computer Terminology

**Transrouter** - The decision support tool for translation managers.

**EUROMAP** - Promoting greater awareness of language technology in the emerging Information and Communication Society.

**DIET** - Diagnostic and Evaluation Tools for Natural Language Applications.

- Other projects proposed to the Commission of the European Communities under the E-Content, the 5th Framework and the Adapt Programmes.

**Next generation localization centres** new project- EU funded project (approx. 16 million Euro’s (Approx. 16 million Euros, 5 years duration,
25% contribution from industry partners in form of manpower, tools and technologies, linguistic resources & joint IPR sharing) with participation from
- Dublin Centre University text analysis, automated text tagging, machine translation (example/statistical based)
- University College Dublin speech technologies
- Trinity College systems framework, content management system
- University of Limerick - localization
- IBM

New tie-ups by Localization Research Centre, University of Limerick

Localization Research Centre (LRC) - Malaysia
The leading Malaysian University, the Universiti Malaysia Sarawak at Kuching (UNIMAS), and the University of Limerick's Localization Research Centre (LRC) establish closer links between the two institutions to promote internationalization and localization related research and teaching.

Localization Research Centre (LRC) South Africa
South Africa's largest university and one of the largest distance universities in the world, the University of South Africa (Unisa), and the University of Limerick's Localization Research Centre (LRC) announce their intention to establish closer links between the two institutions to promote internationalization and localization related research and teaching.

The two organizations are planning joint research projects, staff exchanges and joint funding applications. In addition, the agreement covers the establishment of a mirror site of the LRC's Localization Technology Laboratory and Showcase (LOTS), the holding of LRC South Africa events and training courses, and the initiation of joint research projects in the area of internationalization and localization.

Google look to UL in search for i18n Project Manager
“The University of Limerick has been highly recommended as the premier academic institution that teaches the best minds in the localization / internationalization business,” according to Cyndy Cartwright of Korn/Ferry International, who is working closely with Google to find a project manager to internationalize all of Google's products globally.

Localization Research Centre establishes link with leading Brazilian University Centre GeNESS (UFSC)
The agreement covers, among other points, the establishment of a mirror site of the LRC's Localization Technology Laboratory and Showcase (LOTS), the holding of major LRC Brasil events and training courses, and the initiation of joint research projects in the area of internationalization and localization.

“This memorandum of understanding lays the foundation for the cooperation between Europe's and Latin America's leading research centres. Combining GeNESS's expertise in Latin America's emerging export markets with the LRC's long-standing internationalization and localization research agenda will generate terrific opportunities for both organisations”, said Reinhard Schäfer, Director of the LRC at the University of Limerick.

Research Programme at Localization Research Centre
The LRC carries out research and development in a number of areas of software localisation, including supporting technologies, such as language engineering. University of Limerick awards PhD in localization Research. Following are some of the Research projects undertaken by students for their Masters / PhD degree.

1. The development of an Open Source Localization Package for the Localization of Open Source Software
   Researcher: Kevin Bangry
   Programme of Study: Masters Degree in Computer Science from the University of Limerick
   Status: On-going

2. An Investigation into Aspects of Cultural Theory and their Relevance in the Development of Localized Web Applications
   Researcher: Patrice Fanning
   Programme of Study: Masters Degree in Computer Science from the University of Limerick
   Status: On-going, near completion

   Researcher: Rafael Guzman
   Programme of Study: Masters Degree in Internet Systems from Dublin City University
   Status: Completed

4. Traditional Chinese Internationalization Issues and Their Resolution Using ASP.NET
   Researcher: Joanne Cheung
   Programme of Study: Masters Degree in Software Localization from the University of Limerick
   Status: Completed

5. The development of an XLIFF Source Converter
   Researcher: Kevin Bangry
   Programme of Study: Bachelor of Science Degree in Computer Science from the University of Limerick
   Status: Completed
6. An Approach to Localizing an Existing Website Area

Researcher: Patrice Fanning
Programme of Study: Bachelor of Arts Degree in Languages with Computing from the University of Limerick
Status: Complete

Training initiatives at Localization Research Centre
A. Certified Localization Professional - TILP-supported certification programme
B. Graduate Diploma in Localization Technology
C. MSc in Global Computing and Localization
D. Professional Development Courses
E. LRC Internationalization and Localization Summer School
F. LRC Reader

I) Institute of localization professionals (TILP) -
Certified Localization Professional: The primary aim is to develop professional practices in localization globally. TILP is a non-profit organization, owned by its members and led by a council elected at its annual general meeting.
TILP represents localization industry professionals and professionals active in localization related areas. These include:
1. Software publishers and publishers of other material using electronic media
2. e-content providers
3. localization service providers
4. tools developers
5. trainers and educators (including third level colleges)
6. Researchers
TILP’s founder sponsors include Oracle, Symantec, Alchemy, Microsoft, Novell, LitONBridge, and VeriTest and so on.
The institute of localization professionals runs Certified Localization professional (CLP) certification courses

The Certified Localization Professional (CLP) project was originally funded by the European Union under its ADAPT initiative (1998-2000) and developed a framework for the certification of localization professionals under the leadership of the Localization Research Centre (LRC).
- 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 course and training providers

TILP- CLP broad course structure is as follows
1. The course has contents from three different streams such as linguistic, engineering and management. All these three streams are much relevant to the process of localization.
2. Various levels of courses are offered: level-1, level-2, level-3 and all these level courses has contents from all the three streams as mentioned above
3. Level-1 is the introductory course and part of it is offered online and part of it is offered onsite.
4. The onsite course is delivered through local partner with local contents and local infrastructure.
5. The course duration for level-1 can be roughly 2 months.

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

Certification:
- TILP certifies individuals as having successfully completed a localization course provided by a TILP accredited organization.
- CLP is a part requirement for the admission as Professional member.
- Course providers will pay fees for audits and certificates issued to participants. Fees will include one year associate membership.

Certified localization professional CLP
- Level 1 (introductory)
- Level 2 (Diploma in localization)
- Level 3 (masters programme practicing/experienced people)

Criteria for CLP: localization is multidisciplinary subject consisting of Linguistic, engineering, management...
<table>
<thead>
<tr>
<th>Various Disciplines</th>
<th>TILP - CLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineering</td>
<td>Level - 1 - introductory level courses</td>
</tr>
<tr>
<td>Management</td>
<td>Level - 2 - medium level courses (equivalent to Graduate programme of UL)</td>
</tr>
<tr>
<td>Linguistics</td>
<td>Level - 3 - Advanced courses (equivalent to PhD programme of UL)</td>
</tr>
</tbody>
</table>

1. Course content getting validated and finalized by Jan 2008
2. Online Course start date around April – May 2008
3. On site by June 2008

Courses & Short Term Training Programmes offered by University of Limerick

The University of Limerick (UL) is offering two new programmes in Localization, designed with the support and help of the most eminent industrial and academic experts in the field. The Graduate Diploma in Localization Technology is aimed at those who want to learn how the world's digital publishers localize their products, while the MSc in Global Computing and Localization deals with the underlying scientific and business issues in localization. Each programme is now also available part-time (one-day per week over two academic years).

1) Graduate Diploma in Localization Technology

Graduate Diploma in Localization Technology is a one-year taught programme. It is aimed specifically at students who might not have a technical background but who would like to become involved in localization and, above all, the technical aspects of localization. This programme is aimed specifically at students who might not have a technical background but who would like to become involved in localization and, above all, the technical aspects of localization.

Objectives

- Has a strong focus on Localization Tools and Technologies;
- Looks more closely at Internationalization, moving localization up the value chain;
- Build on the University’s world-wide reputation as the premier academic institution that teaches the best minds in the localization / internationalization business;
- Provide individuals interested in acquiring postgraduate academic qualifications in the areas of localization and internationalization technology with an easily accessible programme to do so.
- Provide an entry point for students aiming at MSc and PhD level qualifications in the areas of localization and internationalization.

2) MSc in Global Computing and Localization

MSc in Global Computing and Localization is a one-year full-time taught programme. It is aimed specifically at students who already have a background in either localization, computing, language technology, translation or related disciplines. It guides students in their research of the underlying issues in localization, with an emphasis on technical and business aspects. Students are encouraged to spend some time studying at one of UL’s partner universities and work closely with industrial mentors on relevant research projects.

Objectives

This programme was prepared in consultation with a large panel of academic and industry experts, and takes into account the changes in localization and internationalization practice. It caters for new and emerging requirements with a focus on high-quality research output. In addition, while it is centred around the technical aspects of localization, it recognizes the need for localization professionals to have a good understanding of international business organization. In particular, the programme now puts a much stronger emphasis then before on:

- Internationalization requirements and their implementation;
- The critical analysis and research of translation technologies;
- The automation of the localization process in what has been called the Localization Factory;
- Business organization and international business practices;
- Build on the University’s world-wide reputation as the premier academic institution that teaches the best minds in the localization / internationalization business;
- Provide individuals interested in acquiring postgraduate academic qualifications in the areas of localization and internationalization with an easily accessible programme to do so, in close cooperation with industrial partners;
- Guide and support students conducting independent postgraduate research in technical aspects of localization / internationalization taking into account the relevant business processes;
- Prepare students for further PhD studies.

3) Professional Development Courses

These courses will be recognized as partial fulfillment of the requirements for the Certified Localization Professional (CLP), grade 1, awarded by the Institute of Localization Professionals (TILP). All courses will be run by professional tutors. Courses will be ‘hands-on’. The maximum number of participants will be 20. Courses will be delivered in state-of-the-art laboratories at the Department of Computer Science and Information Systems, University of Limerick, Ireland.
On successful completion of the course, participants will receive a certificate which will be recognized by The Institute of Localization Professionals as partial fulfillment of the requirements for the Certified Localization Professional (CLP), Grade 1.

**Localization Engineering**
This course provides an introduction to Localization Engineering using Alchemy Catalyst case studies. Successful participants will be recognized as Certified Alchemy Catalyst Users (basic and advanced levels).

**Localization Project Management**
This course provides an introduction to Localization Project Management using case studies.

**Localization Translation and Documentation Engineering**
This course provides an introduction to Localization Translation and Documentation Engineering using TRADOS case studies. Successful participants will be recognized as Certified TRADOS Users (basic and advanced levels).

**Localization QA and Testing**
This course provides an introduction to Localization QA and Test Automation Engineering using SilkTest case studies. Successful participants will also be recognized as automated software test developers (specifically for localized application testing).

**E) LRC Internationalization and Localization Summer School**
The Summer School is aimed at individuals that are researching and studying localization and localization related issues. It aims to provide individuals from different disciplines with the opportunity to experience areas beyond their own expertise.

Experience in computer programming is not required, but will need a working knowledge of computers and the windows operating system.

The LRC Internationalization and Localization Summer School will offer attendees an in-depth look at, and hands-on experience of, localization - from basic introductions to advanced concepts, and will include an intensive one day course in .NET localization.

**Localization Technology Laboratory and Showcase (LOTS) at Localization Research Centre**
The LOTS is run by LRC and enables localization professionals and trainers to identify suitable technology faster and more efficiently. Since September 2002, localization professionals can access the LOTS laboratory free of charge. It offers the tools free of cost for non-commercial training purposes and the LOTS laboratory also helps the professionals to identify suitable technology faster and more efficiently and localisation technology developers will find it easier to establish contacts with potential customers.

**The central point for learning and testing**
- With this facility, researchers and developers can experiment with different technologies and operating systems, as well as language and locale settings.
- It even offers sample files, donated from various companies, to aid their research and evaluation.
- LOTS encourage and facilitate high-end research.

**The localization industry showcase**
- Central repository of commercially available tools and showcase to developers for raising awareness of the tools and technologies
- For the industry - make it easy for potential customers and users to access their products, and expand the market for localization tools and technologies.
- Users have easy, hands-on access to tools and technologies for evaluation purposes

**Physical Access**
- The LOTS laboratory is physically accessible to students and any professional involved in the localization industry both in Ireland and overseas

**Remote Access to the LOTS Server - LOTS Online**
- Most of the applications in the LOTS laboratory can be accessed by logging into the LOTS server remotely (i.e. from anywhere in the world).
- This server contains most of the software applications and sample files available in the actual LOTS laboratory.

**Contents**
**Localization Tools CD provided by LRC**
LOTS CD was handed over by Localization Research Centre, University of Limerick, Ireland to Ms. Swaran Lata and Mr. M. D. Kulkarni. A small agreement in regard with usage of the same only for evaluation purpose was signed between C-DAC and LRC.

The LOTS CD consisted of the following localization tools
1. Catalyst 6.0
2. CatCradle - is a web page editor
3. Lingobit Localiser -
4. Project Open
5. WebBudget.
CATALYST:
Localizer Edition is the visual solution of choice for over 80% of the world's largest software development companies. It is used to reduce localization project cycle times, improve quality and speed up delivery of software applications into international software markets.

Translator/Pro Edition, it comes with a five additional Experts:

- **The Leverage Expert** turns your cost of localization into an investment and reduces the engineering and testing impact during a product update or revision.
- Using the CATALYST Leverage Expert you can take translations from previous product revisions and automatically update these into new revisions, easily, automatically and safely.
- **Project statistic reports** are automatically generated to identify and track all changes to your new revisions and these can be used by Project Managers to track the translation savings due to the leverage process.
- The Leverage Expert is enhanced by the addition of the Update Expert which allows the inline replacement and leveraging of individual files within a TTX.
- **The Pseudo Expert** simulates the effects of translation of your software applications or Internet files and so helps the development of locale neutral applications.
- **The Validate Expert** is enhanced with the addition of our Runtime Validation utility, enabling the detection of common localization bugs while your software application is actually running!
- **The Visual Comparison Expert** enables the localizer to accurately and visually determine the scope of change between software revisions in a matter of seconds.

**CatsCradle is a web page editor**

- Fast and easy to use web page editor for professional language translators. Translate whole web-sites without having to worry about page layouts and HTML code.
- Safe all-in-one web site content localization.
- Grabs all the text that requires translating from a web page, puts it into a built in editor for you to translate alongside, and then automatically integrates your translated localized text back into the web page - leaving all the sensitive HTML code untouched.
- You can instantly preview your work in progress in a web browser at the click of a button. There's even a real-time view where you can see the web page evolving as you translate.

- CatsCradle offers full Unicode support so you can easily perform Cyrillic, Greek, Thai, Chinese and Japanese web site localization.
- Designed for translators, CatsCradle is also an ideal tool for webmasters who want to edit the text content of existing web pages. Additionally, built in support for .hhe and .hbb files means it can also be used to translate chm help files.
- Fast and easy to use.
- Extracts all the text that requires translating from a web page, including hidden text, image alt-text etc. Just type your translations alongside each line of text in the table.
- When done, click 'save' and your translated text is automatically put back into the web page.
- Instant preview of original and translated pages in your web browser at any time.
- Built in automatic glossary panel suggests words and phrases while you are typing - with a single key press to accept a suggestion. Helps you keep key phrases and terminology consistent throughout a project.
- Project Catalogue facility provides a single list of all pages in a project with word counts and translation status to help keep track of progress on larger projects.
- Supports .hhe & .hbb help index and contents files - so can be used to translate chm help files also.
- No pre or post processing of files - once translation has finished, you have a complete translated web site.
- Safe: no worrying about fonts, page layout, hyperlinks, html code, or hidden text. It's all taken care of.

**Lingobit Localizer:**

Lingobit Localizer extracts localizable resources from your application and makes it easy to translate, check and preview translation. When translation is ready, Localizer creates localized version of your application. No source code changing is required. Localization Getting Started demo Software Localization Getting started

Localization can be done in-house or delegated to another company. In the latter case, translation tasks are distributed via self-extracting localization kits with a project file (no source-code) and a 'lite' edition of Localizer for translators. Localization with Translators demo Software Localization Getting started

When you release new version of your software, old translations are automatically merged with resources from the next version and you'll only need to translate new and changed content. Translation Re-use demo Software Localization Getting started
Platforms
Lingobit Localizer is a perfect software localization tool that supports localization of several development platforms allowing translator to change text, position and other localizable parameters.
- C++, Win32 and MFC localization
- Borland Delphi & CBuilder localization
- Java localization
- XML
- Text-based files (*.ini, *.cpp, *.txt)

Key Localization Features
Lingobit Localizer simplifies communication and workflow throughout the entire localization process by offering users a unified interface and productivity tools, such as Automated QA, Validation Expert, Translation Memory, etc. Lingobit's project monitoring tools give a localization manager the ability to see what is going on at each stage of the localization process, which ensures accountability, gives clarity and control to efficiently manage localization across several steps.
- Automatic Validation
- Exchange Wizard
- Translation Re-usage
- Pseudo Translate
- Version Control
- Translation Memory
- and more...

Project-open
Web-based ERP/Project Management software for organizations with 2-200 users. It integrates areas such as CRM, sales, project planning, project tracking, collaboration, timesheets, invoicing and payments.
One of the largest open-source based web applications in the world with more than 1,000,000 lines of code. It is used by more than 1000 companies in 25 countries to run their businesses.

Module Overview
- Finance - Tracking and evaluating company results
- Project Management - Internet Project collaboration and management
- Customer Management - A "CRM-Light"
- Supplier Management - Freelancers etc.
- Human Resources Management - Staff Employees and Skills
- Knowledge Management - Supporting knowledge processes
- Data-Warehouse & Business Intelligence - Analyzing and discovering patterns in your business
- Content & Community - Optional modules related to creating online communities and Internet transactions around your business

- Systems Integration & Interfaces - Modules that connect project-open with other systems.

Finance
The aim of the finance module is to provide the company's senior management with a real-time view to all relevant financial information of the company. For this purpose project-open provides a number of specialized modules that cover all important areas of small and medium-sized project organizations:
- Invoicing
- Timesheet management
- Travel costs
- Fixed costs
- Provider costs (via Web interface)
- Cost Center Permissions
- Financial reporting
- Export interfaces to Excel, KHK Kaufmann, ContaPlus, SQL-Ledger and SAP-FI

Project Management
The PM module integrates project-related information from all project modules into "project rooms" or "e-rooms", allowing you to collaborate online with suppliers and customers (Extranet). Sophisticated access permissions allow you to protect your business critical information.
- Projects, subprojects and project task
- Project templates
- Project-related chat rooms
- Project reports and -tracking
- Risk Management
- "Earned Value" project completion tracking
- Gantt scheduling
- Project file storage
- Project discussions
- Project news
- Incident management
- Import interface to MS-Project
- Integration with cost management
- Integration with invoicing

Customer Management
The Customer Management module ("CRM-light") unifies all functionality related to the management of customer relationships.
- Customer contact management
- Integrated customer interaction history
- Online web registration
- Customer tracking
- Customer classification and status engine
- Import interfaces with MS-Outlook and ACT!
Supplier Management
The Supplier Management Module unifies all functionality related to the management of suppliers in project-oriented organizations. The strategic suppliers for this type of businesses are usually freelancers and other types of human resources, so the boundary between HR and supplier modules is blurry.
- Supplier contact management
- Integrated supplier interaction history
- Freelance skill database
- Supplier web invoice tracking
- Supplier quality module

Human Resources Management Module
The Human Resources Management Module mainly deals with the management of staff employees.
- Employee payroll information
- Employee recruitment process
- Employee portraits are shown together with their office

Knowledge Management Module
- Full-text search engine for Intranet, discussions and external files
- "Expert Finder" to localize a domain expert and post questions
- "Knowledge Market" evaluates knowledge supply and use
- Extensible permission management to control the access to business critical knowledge resources

Translation Module
- Translation workflow
- Translation quality
- Translation project status reports
- Integration with invoicing module

Data Warehouse & Business Intelligence
- Multidimensional view to all corporate data
- Predefined multidimensional cubes
- Implementations available for open-source (Mondrian) and commercial (MS-SQL-Server) OLAP servers
- Predefined models for extraction of CRM key performance indicators (in collaboration with Loyalty Matrix)

Content & Community Modules
Project-open is based on the OpenACS community platform, so that all OpenACS modules are also available for project-open. These modules include:
- A powerful Online store
- Classified Ads (second-hand market)
- Public discussion forums
- Webmail

Jabber Chat
Blog & Wiki

Security
- One-Time-Password Authentication
- Build for the Internet with high-security architecture
- Role-based permission management
- Subadministration of users
- All modules are customizable (open source code).
- Secure-HTML / SSL encryption

Systems Integration
- Authentication integration with corporate LDAP server
- "Multi-Company" allows to manage several separated companies with different URLs on a single server
- XML-RPC Interface

Common Characteristics
- Intuitive graphical user interface
- All modules are customizable (open source code).
- Customized (modified) modules can in general be used with the next version of project-open

WebBudget XT: WebBudget XT is a world class software tool that helps language professionals and localization managers to quickly assess and translate the content of a web project.

TRANSLATE
- Support for most common tagged formats, such as HTML, SGML, XML, ASP, JSP, PHP and variations
- Support for UTF-8 encoding.
- Support for most common scripting languages, such as JavaScript, vbscript.
- Code-free text extraction and segmentation.
- Easy-to-use integrated translation memory.
- Advanced fuzzy logic, including auto-assembling feature
- TMX import/export support.

MANAGE
- Comprehensive text analysis in any language, including double-byte languages.
- Fully customizable reporting options.
- Generate accurate text and images quotations in a snap.
- Exclusive SmartCount technology to handle delimiters and special cases.
Other Useful Features:
- Download sites with integrated new Map a site tool.
- Search the site for specific keywords.
- Compile the e-mail addresses found.
- Batch conversion of HTML files into RTF colored files, highlighting the translatable text.
- Easily create and exchange text extraction/budgeting profiles.

The evaluation of these tools will be carried out at C-DAC, GIST lab and detailed report will be generated for future reference.

Discussions with Prof Cannae - Vice President University of Limerick

We had quite interesting discussions with him in regards with taking the activity forward. He gave a brief of what is the current economic situation in Ireland and what measures are being taken by the Government to increase the R&D base. As regards with localization efforts in Ireland he has pointed out that in recent past MNC’s are building their base for localization and Ireland seems to be a viable proposition. Following are certain reasons for the same:

1. In mid 80’s US companies started expanding their business in the European countries and in order to capture the local market started the localization activities in other Latin based languages.
2. US people feel at home in Ireland.
3. Irish Government gave its focus to localization by giving tax incentives. (Corporation tax was reduced from 40% to 10% than rest of the industry). Hence for cost effective localization Ireland was the destination for most of the MNC’s.
4. Because of lack of industry and opportunities most the Irish people were attracted towards other nearby countries and as a consequence, the population in Ireland saw a dip from 7.5 million to 3.5 million. However in recent past, because of Government measures the same has now increased to 4.0 4.5 million.

Irish Government wants to give more focus on Research (especially PhD level work) rather than shifting the base to the manufacturing industry. (Since the manufacturing industry requires major infrastructure and cheap manpower.)
Discussions with Microsoft's Martin Orsted on localization “Challenges and issues”

Martin Orsted: Senior Group Manager, International product engineering, Ireland

He gave brief introduction of the localization activities happening in Microsoft, Dublin. He touched upon various aspects of localization and informed that he is responsible for localization of Microsoft Office Suite in almost 30 different languages & his team is poised to undertake localization for 100+ languages of world. This includes 8 Indian languages as well.

He also touched upon various requirements of localization and tools being developed in-house by Microsoft for the same. He debated the current standard of XLIFF and indicated that it does not suit the MS localization requirements. They have developed LCX, a similar standard for the same.

He also talked about various skill sets required for linguistic, engineering and management discipline at the skill levels to perform various tasks of localization (such as translation, validation, functional as well as linguistic testing, management and so on).

He gave example of waste of resources because of political issues in the given country for which localization was being undertaken.

It is expected that the localized version of the Office suite needs to be immediately released within 2-3 days after release of English version in US for Spanish, German, French & Dutch languages.

He informed that in order to undertake localization for new language MS expects a minimum installed user base of 3 million licenses.

He also informed that MS has undertaken localization at the instance of the Government in some countries (who has paid money to them) for the languages not covered by MS. MS looks forward to Governments for endorsement of the terminology used for localization.

Following aspects need to be looked upon while gearing up for the localization.

A) User of more controlled English (careful usage of verbs, no long sentences, simple sentences)

B) Linguistic reviews
   a. Bullet proofing source language
   b. Controlled English
   c. Identify key terms
   d. Localize in context
   e. Protect tags
   f. Remove true repeats
   g. Recycle previous translations
   h. Glossaries

C) Some issues with localization
   a. Hard coded strings
   b. String dependencies

c. String length limitations

d. Character corruption

e. Illegal characters

f. Concatenating strings

g. Undocumented placeholders

h. Static dialogs and poor DAL implementation

i. Don’t forget the regional settings

j. Pseudo localization
   i. Adding randomness to pseudo localization
   ii. Hybrid pseudo

k. Approx. 3000 localization files/languages for Microsoft Office products

l. 30 language localization go parallel with bi-weekly updates

m. Move to automation and standards based on XML

n. Microsoft standard LCX similar to XLIFF
   i. Transforming data
   ii. Noise reduction, repeats from component based localization to resource based localization
   iii. Working in parallel, localizing and bug fixing the same file at the same time
   iv. Multiple locator working on the same file
   v. Localization outsourced – file management in-house

o. Components of localization system -
   i. Track changes
   ii. Generate files
   iii. Update files
   iv. Concentrate file management in one place
   v. EAR Eliminate, Automate, reallocate
   vi. Contextual localization
   vii. Localization in cloud
   viii. Crowd sourcing, collaborate approach community effort.

p. Testing - Advanced Test Script Frameworks
   a. Motifs, MAVI (Moving towards AI)

q. Language Interface Pack (LIP) with community glossary

r. EAMT European Association for Machine Translation

s. Linguistic Quality: retains meaning irrespective of age, literacy, region and so on
   i. Combinatorial analysis
   ii. Orthogonal array
   iii. Directed, Ad-hoc, Ad-hoc-Directed

D) Microsoft's 'XP Lite' Microsoft's initiatives in Asia.

- Microsoft's 'XP Lite' - Low-Cost Version of Operating System sold in Asia

- The Starter Edition will ship on new, low-cost desktop PCs in Thailand, Malaysia and Indonesia. Two more countries in the five-country pilot program would be announced later this year following further discussions with governments with special pricing
The key features of the new software would be "localized" help features, country-specific wallpapers and screensavers, and "preconfigured settings" for features that might confuse novices.

**Localization Standards at International levels**

Major localization vendors, backed by the Localization Industry Standards Association (LISA) and The Organization for the Advancement of Structured Information Standards (OASIS), have agreed on open XML based standards for storage and exchange of data in the localization process:

- **Translation Memory eXchange (TMX)** file format for exchanging translation memory data.
- **TermBase eXchange (TBX)** format for terminology exchange.
- **XML Localization Interchange File Format (XLIFF)** for extracting and storing localisable data in a common file format.
- **Open source applications** use GNU Gettext and the Portable Object (PO) file format.
- **SRX 1.0 Specifications - Segmentation Rules eXchange format (SRX)**.
  a. The purpose of the SRX format is to provide a standard method to describe segmentation rules that are being exchanged among tools and/or translation vendors.
- **TR29**
  a. Text Boundaries - describes guidelines for determining default boundaries between certain significant text elements: grapheme clusters ("user-perceived characters"), words, and sentences. For line break boundaries, see Unicode Standard Annex #14 (UAX #14), "Line Breaking Properties."
- **GMX_V1.0 Global Information Management Metrics eXchange (GMX)**
  a. GMX-V adopted as an OSCAR standard for the globalization industry on February 26, 2007.
  b. GMX-V provides a variety of statistics related to word and character counts that can be used to precisely quantify the amount of text (of various types) in a document. While it was designed with localization tasks in mind, it may be used in any field where precise, standardized quantification of text is needed.
  c. The three components of GMX are:
    i. **Volume (V)**
    1. Global Information Management Metrics Volume addresses the issue of quantifying the workload for a given localization or translation task. This is often commonly referred to as word counts. Word counts, however, do not convey the true range of possible statistics that can be used to assess the cost of localizing a document. Global Information Management Metrics Volume is a more precise definition of the statistics necessary for translation billing and sizing purposes.

- **Complexity (C) (proposed)**
  1. GMX-C will quantify the complexity of translation tasks. This format has not yet been defined.

- **Quality (Q) (proposed)**
  1. GMX-Q will specify the quality requirements for translation tasks. This format has not yet been defined.

**H) Internationalization Tag Set - ITS 1.0**

a. Defines data categories and their implementation as a set of elements and attributes called the Internationalization Tag Set (ITS). ITS is designed to be used with schemas to support the internationalization and localization of schemas and documents. An implementation is provided for three schema languages: XML DTD, XML Schema and RELAX NG.

**Presentation by Ms. Swaran Lata**

Ms. Swaran Lata touched upon various activities undertaken by TDIL Programme in giving boost to the Indian language technology development. The presentation covered an overview of activities carried out by TDIL since its inception, new activities carried out in consortia mode, various projects which are funded by TDIL, standardization activity, participation in international standards such as Unicode, W3C, IETF, ICANN, etc.

Her presentation also covered:

- Challenges involved in Localization for the 22 officially recognized languages with diverse scripts and more than one language based on a script & languages having more than one scripts.
- The expectations from this particular visit, especially to explore possible tie up with Localization Research Centre, University of Limerick.
- The objectives of the new proposed centre “National Localization Research & Resource Centre.”
- Complexity of supporting Indian Languages w.r.t. Input mechanisms, rendering issues etc.
- Need to establish languages with other countries for Inter-counters common languages such as Bengali, Tamil, and Nepali etc.
- The National Roll-out Initiative to promote localization by bringing out CDs. Contains BIPK in the 22 Indian Languages.
- The interface with industry which led to availability of Operating systems, Databases, applications in Indian Languages.
- To work closely with e-governance currently under implementation in the country.

Presentation by M.D. Kulkarni

Shri M.D. Kulkarni covered language technology Research & Development activities undertaken at GIST labs of C-DAC, Pune. He also touched upon the importance and involvement for standards development, linguistic diversities, complexities in Indian languages (script and languages), various tools and technologies available for common man through the launch of free software and fonts CD.

He has also showcased the localization framework being developed and deployed for various applications such as Railways, Banking, Insurance, e-Governance and so on.

He opined a bi-lateral co-operation with LRC of University of Limerick for setting up a similar localization Centre and LOTS laboratory in India which will boost the language computing and catalyst the process of content generation in Indian languages.

He also stressed up the point that unless otherwise there are contents in Indian languages, it will be difficult to bring out technologies such as spellcheckers, grammar checkers, thesaurus, Machine Translation system and so on.

Contributors to the LOTS Satellite Distribution
- www.alchemysoftware.ie - Alchemy Catalyst (visual localization solution)
- www.stormdance.net - CatsCradle (web site localization tool)
- www.passolo.com - PASSOLO (visual localization solution)
- www.project-open.com - Project Open (project management and workflow)
- www sdl.com - SDL & TRADOS tools suite (extended LOTS Satellite Distribution only; visual Localization solution and translation memory)
- www.webbudget.com - WebBudget (web site localization tool)

Important Websites/Resources:
1. www.gilc.info/limerick/declaration.pdf - GILC Global initiative for local computing
2. www.electroline.org
3. LOTS online
4. www.igniteweb.org - The IGNITE project is an EU-funded project that began in April 2005 and will run until April 2007. IGNITE will pool together linguistic infrastructure resources and provide convenient access and a market place for them.
5. Localization Focus tools review available in “localization focus” and “multilingual computing”
6. www.ethnologue.com
8. www.localization.ie - The Localization Research Centre (LRC)
9. www.tilopenline.org - The Institute of Localization Professionals (TILP)
11. www.lisa.org - The Localization Industry Standards Association
12. www.unicode.org - The international character encoding consortium
13. www.oasis-open.org - The XML consortium developing open standards (XLIFF, Trans-WS)
14. www.eel-online.org - eContent project

One Laptop per Child - Nicholas Negroponte

Founder and chairman of the One Laptop per Child non-profit association. He is currently on leave from MIT, where he was co-founder and director of the MIT Media Laboratory, and the Jerome B. Wiesner Professor of Media Technology.

A graduate of MIT, Nicholas was a pioneer in the field of computer-aided design, and has been a member of the MIT faculty since 1966. Conceived in 1980, the Media Laboratory opened its doors in 1985. He is author of the 1995 best seller, Being Digital, which has been translated into more than 40 languages.

In the private sector, Nicholas serves on the board of directors for Motorola, Inc. and as general partner in a venture capital firm specializing in digital technologies for information and entertainment. He has provided start-up funds for more than 40 companies, including Wired magazine.
Localization will get boost through the above scheme as there will a huge demand for localized applications, tools and even localized contents.

General discussions with University of Limerick

The localization Research Centre, University of Limerick has given the “Associate membership” for one year free of cost to both Mr. Mahesh Kulkarni and Ms. Swaran Lata

Session on “Introduction to localization” by Prof. Reinhard Schaler which was more focused towards the L10n testing and quality assurance.

- Testing needs to be done on three different levels such as
  - Linguistic (cultural issues, translations, locales, etc)
  - Layout (cosmetic appearance) especially for various dialog, menu items and various resolution devices
  - Functional - critical requirement for web pages localization, functionality to be retained under different operating environments and browsers.
  - Test management also is an important aspect of the testing
  - Testing engineers and the localization engineer's needs to work hand in hand otherwise effective localization will not be possible.
  - The testing time required for localized product is multiple of languages in which it is localized. However, expectation is that the localized product testing should be done in shortest span of time.
  - Testing also needs to be done in real time environment.
  - Testing needs to be automated, there are various tools available for the same.

Four major areas of possible collaborative efforts with Localization Research Centre and the newly proposed centre in India:

- Tools development / distribution: adapting localization tools for Indian languages & distribution rights at lower cost.
- Standardization: access to standards / participation in localization standards development such as LISA, OASIS etc.
- Awareness: publications, workshops, conferences
- Joint PHD programme with University of Limerick
- Certified Localization Professional Training collaboratively with LRC and TILP.

Objectives of the Proposed National Localization Research and Resource Centre (NLRRC)

The NLRRC is proposed to be set up in two phases:

Phase I Implementation through the project for initiating activity and also for devising a suitable model of the proposed Centre including draft MOU between NLRRC and LRC, Ireland in above areas. Also planning to put up physical infrastructure for the Centre and obtaining necessary Govt. approvals. Project phase duration may be 18 months.

Phase II: the actual implementation of the NLRRC. The objectives of the proposed NLRRC are:

A) Standards & Certification

Phase I

1. Consolidation & Integration of existing activities of W3C Indian office.
2. Activities covering following standards:
   - W3C
   - Storage
   - Font
   - Keyboard
   - Rendering engines
   - Transliteration
   - Ground work for certification
   - Membership of various localization standardization bodies.

Phase II

- Advanced W3C activities
- Deploying certification scheme for language technology solutions.

B) Localization Tools & Technology Development and Demonstration

Phase I

- Identification of various tools for localization, evaluation, distribution and licensing policies

Phase II

- Setting up of LOTS

C) Technology Support to the application providers

Phase I

- Best practices, guidelines, testing methodologies to be made available on the web.
- Awareness programmes, workshops, conferences on internationalization and localization.
- Development of standardized templates and initiation of linguistic resource creation.

Phase II

- Linguistic resources, tools and tools training, publications
D) Training on localization at various levels

- Adaptation of course curriculum as per Indian language localization requirements
- Initiation of Level I CLP Programme

Phase II
- Post Graduate Programmes and joint PHD Programmes.

E) Consultancy, Education & Outreach

Phase I
- Deploying web portal inclusive of showcase of tools and technologies and integration of ILDC.

Phase II
- Case studies, consultancy and outreach will be implemented in Phase II.

F) Contribution to common locale data repository (CLDR)

Phase I
- To evolve mechanism for participation in CLDR by collaborating with State Govts. and CIIL, Mysore.

Phase II
- Development of vetting mechanisms before finalizing contributions to CLDR.

G) Linguistic Resource Development & Dissemination

Phase I
- Identification of linguistic resources and developing standards, frameworks, guidelines for Linguistic Resource Development.

Phase II
- Building of Resources, validation and mechanisms for distribution

H) E-Content proliferation in Indian Languages

Phase I
- Content Development and management tools

Phase II
- Hosting of content and developing distributed models of content generation.

Courtesy:
Ms. Swaran Lata
Director-TDIL Programme
MC&IT, Govt. of India
slata@mit.gov.in

Mahesh D Kulkarni
Programme Coordinator
C-DAS, GIST, Pune
mdk@cdac.in