Hotel cut-off:
09/30/2009

Venue:

Hilton San Jose
300 Almaden Blvd.
San Jose, CA 95110

Program - Session Descriptions

Wednesday, October 14, 2009

09:00-12:30	MORNING TUTORIALS

Presenter:

Richard Ishida
Internationalization Lead, W3C

Track 1: An Introduction to Writing Systems & Unicode

The tutorial will provide you with a good understanding of the many unique characteristics of non-Latin writing systems, and illustrate the problems involved in implementing such scripts in products. It does not provide detailed coding advice, but does provide the essential background information you need to understand the fundamental issues related to Unicode deployment, across a wide range of scripts. It has also proved to be an excellent orientation for newcomers to the conference, providing the background needed to assist understanding of the other talks! The tutorial goes beyond encoding issues to discuss characteristics related to input of ideographs, combining characters, context-dependent shape variation, text direction, vowel signs, ligatures, punctuation, wrapping and editing, font issues, sorting and indexing, keyboards, and more. The concepts are introduced through the use of examples from Chinese, Japanese, Korean, Arabic, Hebrew, Thai, Hindi/Tamil, Russian and Greek. While the tutorial is perfectly accessible to beginners, it has also attracted very good reviews from people at an intermediate and advanced level, due to the breadth of scripts discussed. No prior knowledge is needed.

Presenter:

Addison Phillips
Globalization Architect Lab126 (Amazon)

Track 2: Internationalization: An Introduction, Part I: Characters and Character Encodings

What is internationalization? What do developers, product managers, or quality engineers need to know about it? How does a software development organization incorporate internationalization into the design, implementation, and delivery of an application?

This tutorial track provides an introduction to the topics of internationalization, localization and globalization. Attendees will understand the overall concepts and approach necessary to analyze a product for internationalization issues, develop a design or approach, and deliver a global-ready solution. The focus is on architectural approaches and general concepts, but will include specific examples and exercises.

Part I focuses on characters, character encodings, and the basics of Unicode.

Presenter:

Elizabeth Pyatt
Instructional Designer Penn State

Track 3: Building a Custom Keyboard Layout for the Mac with Ukulele and XML

Building custom keyboards can be a useful timesaver if you work with an unusual range of characters across a large number of documents. The tutorial will describe how to create a custom keyboard layout on the Mac OS X platform using the freeware Ukelele tool from SIL plus modifications to the XML file. Although the main example will a keyboard built for symbolic logic characters, the tutorial will cover how to create keyboards for many foreign languages.

10:30-10:45 - Morning Refreshments

Presenter:

Addison Phillips
Globalization Architect Lab126 (Amazon)

Track 2: Internationalization: An Introduction, Part II: Writing Global-Ready Code

Part II focuses on preparing for the localization (translation) of user interfaces; making applications "locale-aware", including format and display differences; as well as approaches to delivering multi-lingual and multi-locale software or content.

Presenter:

Thomas Milo
President DecoType

Track 3: Arabic Script: Structure, Geographic and Regional Classification

A new tutorial about Arabic script (including Arabic script for dummies, structural analysis, typology, stylistic geography, technical and aesthetic aspects, language-dependant preferences within calligraphic styles, and extra attention for orthographies East of Iraq), against the background of the development of a brand-new Nastaliq typeface that covers the Unicode for all languages that require this Persian-derived style.

12:30-13:30 - LUNCH

13:30-15:30	AFTERNOON TUTORIALS

Presenters:

Craig Cummings
Mike McKenna
Internationalization Architects
Yahoo! Inc.

Track 1 - Unicode - A Grand Tour
This tutorial will cover the next level of detail of what Unicode is, and how it is used in the real world. The modules of the tutorial will cover: The Unicode standard - what are the "Guiding Lights", or design principles behind Unicode? A tour of Unicode's structure, encoding forms, behavior, technical reports, database, and how to use the Unicode Standard. Implementation according to Unicode - a walk through the details of attributes, compatibility, non-spacing characters, directionality, normalization, graphemes, complex scripts, surrogates, collation, regular expressions and other aspects according to the Unicode Standard and associated Technical Reports. Unicode and the Real World - an overview of International Components for Unicode (ICU) and implementations supporting Unicode in web servers, application servers, browsers, C/C++, Java, PHP, SQL, and various operating systems. On-going programs - how Unicode is evolving to support more minority scripts, languages, and help solve linguistic processing issues.

Presenter:

Tex Texin
Xen Master
XenCraft

Track 2 - Web Internationalization - Standards and Best Practices
This tutorial is an introduction to internationalization on the World Wide Web. The audience will learn about the standards that provide for global interoperability and come away with an understanding of how to work with multilingual data on the Web. Character representation and the Unicode-based Reference Processing Model are described in detail. HTML, XHTML, XML (eXtensible Markup Language; for general markup), and CSS (Cascading Style Sheets; for styling information) are given particular emphasis. The tutorial addresses language identification and selection, character encoding models and negotiation, text presentation features, and more. The design and implementation of multilingual Web sites and localization considerations are also introduced.

Presenter:

Jim DeLaHunt
Principal
Jim DeLaHunt & Associates

Track 3 - Building Multilingual Websites in Joomla [Drupal]
A practical look at the language and locale capabilities of Joomla! and Drupal, two leading free software content management systems (CMSs). They let you build more powerful, more international websites faster. We look at: their core services for internationalization and locale support; localization of UI and content; and localization support in some leading modules. You will leave with specific tips for building your own site. We don't assume Joomla or Drupal experience, but do include material for advanced practioners. A good tutorial for web site product managers, for web designers and developers, and for managers of international web site teams.

15:30-15:45 - Afternoon Refreshments

15:45-17:45	AFTERNOON TUTORIALS

Track 1 - Unicode - A Grand Tour (Cont'd.)

Presenter:

Richard Ishida
Internationalization Lead
W3C

Track 2 - Creating XHTML/HTML Pages with Right-to-Left Scripts
This short tutorial explains how to go about creating XHTML and HTML pages containing text written in the Arabic or Hebrew scripts. The tutorial examines how best to achieve the correct effect for these bi-directional scripts using appropriate markup, CSS properties and Unicode code points or entities. It covers the basics, and goes beyond to provide recommended techniques for some of the tricky situations that even native speakers can struggle with. The tutorial assumes a basic familiarity with the bi-directional characteristics of Arabic and Hebrew, as well as a basic knowledge of HTML and CSS.

Presenter:

Behdad Esfahbod
Software Developer
Red Hat/GNOME

Track 3 - Free Software Stack for Unicode Text Rendering
The Free Software world has a lot to offer when it comes to building a stack up from the grounds. Be it building an ARM-based Linux mobile platform or cross-platform text rendering to rendering downloadable CFF fonts on Windows, the Free Software stack provides all the bits and pieces one needs to assemble a high quality OpenType-based Unicode text rendering pipeline with great flexibility. In this tutorial we will go over the building blocks involved and how to put them together.

18:00-19:00 - Welcome Reception hosted by Adobe Systems

Thursday, October 15, 2009

09:00-09:15	*WELCOME & OPENING REMARKS*

<table>
<tr>
<td>09:15-10:00

<u><i>Nicholas Ostler</i></u>
<i>Chairman
Foundation Endangered
Languages</i></td>
<td><i>KEYNOTE Presentation: The Alphabetic Principle and its Enemies</i>
The alphabetic principle for writing seems brilliantly simple, and its implementation, often subverting other options, has often caused explosive growths in literacy, with important historical consequences for cultural survival. Its great advantages are economy of effort in the learner, and ready application to new languages. However, it has drawbacks as to speed for the initiated user, and also (by being essentially mechanical and phonetic) in representing many of the cultural overtones which people like their written language to have. There is, too, a certain resistance to the role of art in writing. But as alphabetic traditions age, becoming less purely alphabetic, these disadvantages can be reduced. New structures may emerge, meaningful patterns that leave alphabets far behind. Alphabetic scripts have more recently revealed new aspects, defining a convenient order to index anything, inspiring the phonemic principle of structural linguistics, and later mapping more easily than other systems onto digital systems, and hence a whole new set of functions for written language. But the alphabet remains a rather arbitrary means of representing meanings, since its icons are parasitic on the particular sounds of particular words in particular languages, a long way from thoughts.</td>
</tr>
</table>

10:00-20:00 - EXHIBIT AREA OPEN

10:00-10:30 - Morning Refreshments in Exhibit Area

10:30-11:20	**SESSION 1**

Presenter:

Kirti Velankar
*Senior Software Engineer
Yahoo! Inc.*

Track 1 - Internationalization with PHP
PHP is one of the most prominent and popular platforms for modern Web development. This updated session discusses PHP from the perspective of internationalization, what some of the challenges in PHP are, the features available in PHP 5, and the promise of Unicode in PHP 6.

This session also includes examples and usage in practical scenarios. You will learn how to effectively build applications for multiple languages and cultures using PHP with some of the new internationalization features such as locales, sorting, resource bundles, as well as date, number and message formatting.

Presenter:

Ken Lunde
*Senior Computer Scientist
Adobe*

Track 2 - Designing & Developing Pan-CJK Fonts for Today
Designing and developing Pan-CJK fonts, meaning fonts whose CJK Unified Ideographs can serve more than a single CJK locale, region, or culture, is both challenging and time-consuming. But, like most things that require effort, there are great rewards: smaller overall font footprint, design consistency across locales, and so on. In developing such fonts, there are challenges related to the actual design of the glyphs, which transcend any font format concerns. This presentation pinpoints specific design and implementation problems that developers of such fonts will face, and then details workable solutions. A prototype Pan-CJK font will demonstrated during the presentation.

Presenter:

Mark Davis
*Sr. Internationalization
Architect
Google Inc.*

Track 3 - Unicode Update: Unicode 5.2 and CLDR 1.7
The 5.2 version of Unicode (Fall 09) adds many new characters, new properties, and fixes to existing properties, and is being issued as a complete online book. CLDR 1.7 (Spring 09) contains over 21% more locale data than the previous release, with over 40,000 new or modified data items from over 140 different contributors, including Adobe, Apple, Google, IBM, and Sun, plus official representatives from a number of countries.

This presentation, from the president and co-founder of the Unicode consortium, covers the new features of both standards, examples of the impact on companies such as Google, and future directions for these and other globalization standards -- the new emoji characters, international domain names, Unicode security, and others.

11:30-12:20	**SESSION 2**

Presenter:

Martin Duerst
Aoyama Gakuin University

Track 1 - Internationalization in Ruby 1.9
Ruby is a purely object-oriented scripting language which is easy to learn for beginners and highly appreciated by experts for its productivity and depth. Internationalization of Ruby made a big leap forwards when this January, Ruby 1.9.1, the first stable release of the Ruby 1.9 series, was released. While previous versions of Ruby mostly treated text data as byte sequences, strings in Ruby 1.9 are sequences of characters. Because Ruby tags each string with encoding information internally, different applications can choose different internationalization models.

The presentation will give a short overview of Ruby as a programming language, and introduce the new internationalization features in detail. We will be concentrating on how to use Ruby with Unicode, which in Ruby's case means UTF-8. We will also discuss internationalization support in Ruby on Rails, the popular Web application framework written

in Ruby.

Presenter:

Kamal Mansour
*Manager of Non-Latin
Products
Monotype Imaging*

Track 2 - Unicode & Fonts: a status report
The adoption of Unicode as the universal character code standard has profoundly changed the computing landscape. We now expect to be able to exchange multilingual text documents across platforms and software applications. Since its inception, Unicode has cautiously distanced itself from the process of displaying glyphs, delegating it to an external "rendering layer" that includes fonts. Alongside Unicode, the OpenType Standard has enabled new levels of sophistication in fonts. However, one is often disappointed by a particular font doesn't work as it should. We will give a brief overview of what works today and what we can expect in the future.

Presenters:

Deborah Anderson
*Project Leader, Script
Encoding
Initiative, Department of
Linguistics, UC Berkeley*
Richard Cook
*Post-Doctoral Researcher,
Dept. of Linguistics
UC Berkeley*
Charles Riley
*Catalog Librarian for
African Languages
Yale University*
Anshuman Pandey
*C.Phil. History
University of Michigan*

Track 3 - Patching Holes in the Unicode Pipeline: A Status Report on the Unencoded Scripts of Asia and Africa
In 2002, 96 scripts listed on the Unicode Pipeline were unencoded. Today, the number is considerably smaller. Currently about 25 scripts from Asia and Africa remain unencoded, but they present particular challenges: many are not well-known and will involve considerable research to acquire materials and to track down experts. This session will be made up of 3 speakers who have worked on South Asian and African script proposals. They will discuss the work that remains to be done and highlight specific issues for implementers.

12:30-13:30 - LUNCH

13:30-14:20 | **SESSION 3**

Presenter:

Norbert Lindenberg
*Internationalization
Architect
Yahoo! Inc.*

Track 1 - Internationalization for JavaScript Applications
JavaScript, as defined by the EcmaScript standard and implemented in browsers, is a rather weak platform for internationalized web applications. Several toolkits have attempted to fill the gap in different ways, ranging from reliance on existing server-side internationalization libraries to implementing the functionality in JavaScript itself. This presentation surveys the landscape and compares the different solutions.

Presenter:

Ken Lunde
*Senior Computer Scientist
Adobe*

Track 2 - The Design & Development of Fully Proportional Japanese Fonts
Japanese fonts have traditionally been designed on the principle that each glyph occupies a fixed design space. Some fonts have overcome this principle by providing alternate metrics, which really amount to pseudo proportional metrics. It is possible to develop Japanese fonts whereby each glyph has proportional metrics by default, in both horizontal and vertical writing directions. In addition to the obvious design challenges, there are also several technical hurdles related to implementing the typeface design as an OpenType font. This presentation details the unique design aspects of Kazuraki, a fully-proportional Japanese font, along with details about its OpenType implementation.

Presenter:

Martin Duerst
Aoyama Gakuin University

Track 3 - Update on Internationalized Domain Names and Internationalized Resource Identifiers
In domain names such as www.unicode.org, only a limited number of characters are allowed. This limitation also applies to Uniform Resource Identifiers (URIs) such as http://www.unicode.org. Internationalized Domain Names (IDNs) and Internationalized Resource Identifiers (IRIs) changed this a few years ago, both allowing a wide range of characters from the Unicode repertoire. The specifications underlying these technologies are currently facing an overhaul, major for IDNs and minor for IRIs. The long-overdue and now imminent introduction of the first international top-level domain names will mean that the importance of IDNs and IRIs will significantly increase in the near future.

The presentation will give a general overview of IDNs and IRIs and discuss the current revisions of the specifications in detail. For IDNs, the set of allowed characters is defined using an inclusion-based model rather than the earlier exclusion-based model. Fixed tables are replaced by a property-based selection process to avoid fixing the specification to a single version of Unicode. The mapping step (dealing with casing and normalization, among else) is moved out of the core libraries and closer to the user to allow adaptions for special

cases and reduce user surprises. The IRI specification is being extended with descriptions of widely used variants for handling characters strictly speaking not allowed in IRIs. Both specifications are affected by bug fixes to bidirectionality restrictions.

14:30-15:20	SESSION 4

Presenter:

Umesh Nair
Software Engineer
Google Inc.

Track 1 - Implementing International Calendars in JavaScript
Conversion routines between the Gregorian calendar and non-Gregorian calendars involve complex floating point computations, large lookup tables and calendar-specific computations. Floating point operations impact performance and accuracy, while lookup tables impact memory footprint and download time. Calendar-specific computations require special algorithms and data structures. Implementing such algorithms efficiently with compact data structures is essential for the successful deployment of online calendars for the international audience. This presentation discusses several such techniques for calendrical calculations in client-side JavaScript. The techniques described here are applicable to a number of other areas in internationalization as well as general software usage with JavaScript.

Presenter:

Thomas Milo
President
DecoType

Track 2 - The Unicode-based Koran: a Conflict Between Calligraphic Tradition and Computer Typography

A technical talk about the practical problems encountered in the project to produce a Unicode-based Koran on the behest of the Omani Ministry of Awqaf and Religious Affairs. The focus is on the discrepancies discovered between the age-old calligraphic tradition and the 1924 revision of the Koran. The pivotal issues will be identified and explained. A workable solution will be presented.

Presenters:

Mark Davis
Sr. Internationalization Architect
Google Inc.
Addison Phillips
Globalization Architect
Lab126 (Amazon)

Track 3 - Language Identification and Usage
In 2006, the IETF issued an updated version of BCP 47 "Tags for Identifying Languages", which updated the way languages are identified in most computer programs and protocols. The latest version of BCP 47 (2009) incorporates over 7,000 new languages and many other improvements. This presentation, from the authors of the updated and previous RFCs, covers:

- the format of language tags and the language subtag registry
- the matching algorithms for comparing language tags to user preferences
- plus distance-based algorithms
- the new features in BCP 47 and their impact on developers

and how BCP 47 is being used in:

- Unicode locales (CLDR)
- prominent open-source libraries such as ICU
- companies such as Google and Amazon

15:20-16:00 - Afternoon Refreshments in Exhibit Area

16:00-16:50	SESSION 5

Presenters:

Steven Loomis
Software Engineer
IBM
Markus Scherer
Unicode Software Engineer
Google Inc.

Track 1 - What's New with ICU
The International Components for Unicode library, or ICU, provides a full range of services for Unicode enablement, and is the globalization foundation used by many software packages and operating systems. Freely available as open-source, it provides cross-platform C, C++ and Java APIs, with a thread-safe programming model. This presentation will provide a brief overview of ICU, with emphasis on the current status of ICU (4.2), including the latest support for Unicode 5.1 and CLDR 1.7, and an update on ICU's planned direction for 4.4 and future releases.

Presenters:

Michael Manca
Project Manager and Solution Quality Analyst
IT Flex Services

Track 2 - A Systematic Approach to I18N Testing
Building on last year's presentation "We're World-Ready, What Does This Really Mean?", Intel's localization experts will present and discuss the steps they follow, the tools they use, and their overall I18N testing philosophy. They will explain in details how they proceed when working with development teams to ensure applications are properly internationalized before they're released or localized. Based on recent I18N testing efforts conducted by Intel, this

Intel Corporation
Tomas Galicia
Solutions Quality Analyst
IT Flex Service
Intel Corporation
Loic Dufresne de Virel
Localization Strategist
IT Flex Services
Intel Corporation

interactive session will provide a solid framework of reference for I18N testing, as well as valuable pointers that can be easily and directly applied to your own localization projects or reused within your organization.

Presenter:

Toshiya Suzuki
Research Assistant
Hiroshima University

Track 3 - Investigation of Opaque Glyphs Synthesized from Old Hanzi
After the long efforts during 7 years, finally ISO/IEC 10646:2008 have included CJK Unified Ideographs Extension C. It has 366 glyphs taken from "Index to Collections of the Inscriptions in Yin-Zhou period" (I2CIYZ) proposed by PRC, and more glyphs are scheduled for future Extension E project. They are suspected to be the glyphs invented only for the specification of Old Hanzi. In this report, the source is investigated and compared with existing dictionaries for Bronze scripts. The requirements of some glyph shapes are questionable, the expected procedure to standardize these opaque glyphs is discussed.

17:00-17:50	SESSION 6

Presenter:

Behdad Esfahbod
Software Developer
Red Hat/GNOME

Track 1 - HarfBuzz, the Free and Open OpenType Shaping Engine
In this session we will introduce HarfBuzz, the unified Free Software and Open Source, OpenType-based, text shaping engine. We will discuss design considerations, technical decisions made, and performance and other features that make HarfBuzz an attractive alternative to the existing OpenType engines.

HarfBuzz is already being used by both GNOME and KDE desktop environments and is at the heart of the GTK+ and Qt desktop and mobile platforms, with others planning to use it in the coming months, including Mozilla Firefox, OpenOffice.org, and ICU Layout.

Presenters:

Andrew Swerdlow
Internationalization Tech Program Mng
Google Inc.
Manish Bhargava
Google Inc.
Jens Riegelsberger
Google Inc.
Laura Cuozzo
Google Inc.

Track 2 - Google Internationalization Quality Control Framework
There are many obstacles to a great international user experience. There is a range of issues that cut across organizational boundaries, such as localization, internationalization, visual design, interaction design, business analysis, usability analysis, and market research. Against this backdrop we at Google started experimenting with a standardized review framework that relies on a global network of external evaluators. These evaluators live in market and thus are familiar with local standards and practices. This framework allows us to identify themes that may point to requirements that are common across multiple regions aiding in prioritizing features or giving resources to projects.

Presenter:

Murry Sargent III
Partner Software Design Engineer
Microsoft

Track 3 - Math Editing and Display in Microsoft Office
Math editing is described that uses math context menus, a math ribbon, keyboard navigation, and formula autobuildup in Microsoft Office 2010. The math typography is similar to TeX's, the input methods are state of the art, the math character set is Unicode's, and the environment is Office's, which comes with the many features one expects from a leading office suite. Demonstrations will be given using Office 2010.

18:00-20:00 - IUC32 CONFERENCE RECEPTION *(IN EXHIBIT AREA)*

Friday, October 16, 2009

09:00-09:50	SESSION 7

Presenter:

Douglas Davidson
Software Engineer
Apple, Inc.

Track 1 - International Features of Mac OS X Snow Leopard
From its inception, Mac OS X has been designed with top-to-bottom international and multilingual support. The latest version, Mac OS X 10.6 Snow Leopard, expands on that with new bidirectional input support, multilingual spellchecking, and many other new features. This session covers the international capabilities of Mac OS X from both a user

and a developer perspective, with a particular emphasis on new features in Snow Leopard. Topics covered include localization, locale data, text input, text display, proofing tools, and user customization.

Presenter:

Brent Ramerth
Software Engineer
Apple, Inc.

Track 2 - International Features of iPhone OS
The iPhone OS platform starts with the internationalization architecture fundamental to Mac OS X, and adds a unique virtual keyboard and text input system that handles a wide array of languages. This session covers the international capabilities of the platform from both a user and a developer perspective, with particular attention to iPhone-specific features. Topics covered include localization, text display, and text input.

Presenter:

Elizabeth Pyatt
Instructional Designer
Penn State

Track 3 - Practical "Unicode Logic" for Online Tech Courses
This session describes some of the challenges and workarounds for implementing Unicode content in two online courses in symbolic logic and thermodynamics. Topics include development utilities, templates and guidance for students, issues with multiple applications and font selection across platforms. The presentation will also discuss some differences between implementing Unicode for math courses and Unicode for foreign language courses.

| 10:00-10:50 | SESSION 8 |

Presenter:

Derek Murnam
Senior Program Manager
Microsoft Corporation

Track 1 - Windows 7: Writing World-Ready Applications
This session centers on the new globalization features for Windows 7, including sorting and string comparison, locale support, and coverage for new languages, with an eye to helping developers extend their applications to a global user base. In addition to introducing the Extended Linguistic Services API, this session will also cover the Multilingual User Interface (MUI) resource technology available in Windows 7. This session will provide an end-to-end look at how to make your application world-ready so that you can easily take your application worldwide and extend your customer base into new language markets.

Presenters:

Markus Scherer
Unicode Software Engineer
Google Inc.
Katsuhiko Momoi
Staff Test Engineer & I18n
Consultant
Google Inc.
Mark Davis
Sr. Internationalization
Architect
Google Inc.

Track 2 - Emoji in Unicode: Cell Phones Meet the Internet
Emoji" symbols or "picture characters" are used in email by more than 80 million Japanese cell phone users. They are treated as characters, via vendor-specific extensions of the Japanese character sets. Other email providers have to be able to exchange emails with the Japanese cell phone companies without losing or corrupting data. Most email providers use Unicode, requiring conversion of mail data to/from Unicode. Unicode Private Use characters are used for this purpose. However, they do not provide for reliable public interchange. For a permanent solution, the Unicode Consortium has approved the addition of the Emoji symbols to Unicode 6.0, and is working with ISO to ensure inclusion in the corresponding version of ISO 10646. This paper presents the state and progress of the Unicode encoding proposal with an overview of the Emoji symbols.

Presenter:

Adam Asnes
President
Lingoport, Inc.

Track 3 - Creating an I18n Project Plan
Many initial internationalization scoping efforts focus on creating findings documents. But often the real trick is gathering accurate metrics and turning them into realistic, budge table and actionable project plans. In this presentation we will demonstrate how we assess source code and architecture, and then review a detailed project plan and how we arrived at tasks, durations and staffing.

10:50-11:10 - Morning Refreshments

| 11:10-12:00 | SESSION 9 |

Presenter:

Mihai Nita
Globalization Architect
Adobe Systems, Inc.

Track 1 - Accessing Globalization Services on Multiple Operating Systems
This presentation will cover the experience gained by implementing a cross platform C library that makes use of the operating system dependent services for the following language and region specific functionality. In contrast to ICU which carries its own set of locale data, this solution provides a cross platform set of APIs but uses the facilities provided by the operating system. This presentation will explore the pros and cons of such an approach, trade-offs, implementation issues, major traps, and some of the surprises we encountered.

Presenters:

Loic Dufresne de Virel
Localization Strategist
IT Flex Services
Intel Corporation
Michael Kuperstein
Senior Localization
Engineer
IT Flex Services

Track 2 - Taking Moblin to the World
When the Moblin project asked for our help to localize their application, our initial reaction was enthusiastic! "Finally a cool open-source project to work on", we thought! After getting back to our senses, we realized that localizing Moblin (Moblin stands for Mobile Linux) was not our typical localization project… Far from it! In this session, we will review the thought process we followed to define and limit the scope of this significant undertaking, give an update on the current status of this on-going project, explain how we addressed the first major challenges of this amazing journey, and provide an overview of the first-ever attempt at community-based translation by Intel's localization team.

Intel Corporation
Margie Foster
*Localization Project
Manager
Moblin Project
Intel Corporation*

Presenter:

Cindy Conlin
*Senior Engineer
The Church of Jesus Christ
of Latter-Day Saints*

Track 3 - Building a Global Names System: A Case Study
This case study discusses our experience building a global names application containing records for all members of the LDS Church worldwide. We'll discuss the interesting challenges and requirements we face, such as building a data structure flexible enough to accommodate names from multiple cultures simultaneously. We'll talk about using ICU's transliteration functionality to generate romanizations of non-Latin names, and about our experience supporting private-use characters in Chinese names. We'll also discuss how we've created a user interface that allows users from multiple locales to work with data that originated in many other locales.

12:00-13:00 - LUNCH

13:00-13:50	SESSION 10

Presenter:

Sumit Sarkar
*i18n Product Specialist
DataDirect Technologies*

Track 1 - Internationalization in Database Drivers for C/C++/Java/.NET Applications
Everything you want to know about i18n and database drivers across C/C++/Java/.NET programming languages. Discussion starts by asking what Unicode support encompasses at the Database Access API level, and what components affect Unicode Support. Take a closer look under the covers at the low level data access across major RDBMS including DB2, SQL Server, Oracle, and Sybase. This includes identifying who is doing the conversions at each component of the data access application layer. To summarize and apply the learned concepts, host will answer key questions about your globalized application's data access: Why should conversions be avoided when possible; and what high level features of a database driver are recommended?

Moderators:

Steven Loomis
*Software Engineer
IBM*
Mark Davis
*Sr. Internationalization
Architect
Google Inc.*

Track 2 - Deploying the Common Locale Data Repository (CLDR)
The Common Locale Data Repository is a project for the exchange of language and locale information used in application development, and to gather, store, and make such data publicly available. By pooling resources, the time and expense of collecting good data is minimized, and language groups have an avenue to get their data into implementations. This session will discuss implementation of CLDR, the latest project status, and how the process is being improved to produce higher-quality data. Ample time will be given for comments and questions from the audience.

Presenter:

Chris Weber
Casaba Security

Track 3 - Unicode Transformations and Security Vulnerabilities
Web-applications are being exploited every day as attackers find new vectors for performing cross-site scripting attacks. This talk will cover ways which latent character and string handling can transform clever inputs into malicious outputs. Many application frameworks such as .NET and ICU enable these behaviors without the developer's knowledge. String transformations through best-fit mappings, casing operations, normalization, over-consumption and other means will be discussed, with inputs useful for testing. A testing tool is also planned for release.

The current state of visual spoofing attacks will also be discussed. Phishing attacks are prevalent on the Web, and well-designed URL's can increase an attack's chance of success. It's eye-opening to see demonstrations of just how vulnerable modern Web browsers still are to many forms of visual spoofing attacks.

14:00-14:50	SESSION 11

Presenter:

Su Liu
*AIX Globalization Architect
IBM*

Track 1 - Unicode Technology and Globalization Support in IBM UNIX, AIX
AIX, an IBM UNIX, supports more than 60 languages and about 250 locales. Unicode is a key technology to support globalization features to meet different national language requirements. This presentation discusses Unicode impacts on globalization strategy and mechanism in UNIX operating system level. It focuses on how Unicode technologies are used to simplify globalization configurations first. Then, topics are covered on Unicode impacts on system performance, locale data test, and national language support procedure. Examples are given to explain show Unicode support on complex texts, CJK input methods, Unicode conversions, and automated tests. A further looking into Unicode highlights customization subjects on user-defined locale settings and user-defined Unicode conversion tables. Finally, issues in implementations, market requirements and solutions for future Unicode support in UNIX are assessed.

Presenters:

Track 2 - CLDR on the Cloud

Benedicto Franco Jr.
Software Engineer
Yahoo! Inc.
Marco Aurelio Carvalho
Senior Software Engineer
Yahoo! Inc.

The value of CLDR (Common Locale Data Repository) for global applications is undeniable. But how do you update time zone and daylight saving rules, or a new currency, or geo-political changes that might be relevant for the application without taking the inherent risks and costs of a release deployment process? In this presentation, we are going to talk about a solution that exposes CLDR as a service and how CLDR on the Cloud can be used to help create robust internationalized JavaScript and Ajax applications fed by CLDR data published in JSON format ubiquitously.

Presenter:

Tex Texin
Xen Master
XenCraft

Track 3 - My Unicode Disk Storage Went into the Circular File
This session will present some of the difficulties of providing a common international interface to file services on different operating systems. Although Unicode supports all the necessary characters, identifying the set of characters that are legitimate on any OS can be difficult, and rules for case-insensitivity, normalization, etc. vary, and may even vary by user. The presentation will describe the problem space. It may offer possible solutions.

14:50 – 15:10 - Afternoon Refreshments

15:10 - 16:00   **SESSION 12**

Presenters:

Ryan Cavalcante
Software Development Engineer
Microsoft
James Lyle
Program Manager
Microsoft

Track 1 - Extended Linguistic Services in Windows 7
In this presentation we will discuss the Extended Linguistic Services (ELS) platform, new to Windows 7, which provides diverse linguistic services to developers through a common API. We will discuss the linguistic services now available to developers through the ELS platform in Windows 7—Language Detection, Script Detection, and various Transliteration services—as well as the future vision for the platform.

Presenter:

Adil Allawi
Technical Director
Diwan Software Limited

Track 2 - Mashing-up Bi-Di
Mash-ups is a relatively new fashionable word on the Web - taking bits of other web sites to build up your own web page. It is not new or special - any search engine showing a snippet of a web site that it has found is a form of mash-up. Integrating a news or micro-blogging feed is another. And it seems that every company and their mother has its own mash-up API. But what happens when you have an Arabic web-site integrate content that may be Arabic or English or both? The Unicode Bi-Di Algorithm can render text and numbers unreadable. URL's may become unusable or, in the worst case, direct to fraudulent sites. It can be hard to predict how to mark-up the integrated content for the right result. This presentation will cover real world issues and attempt to suggest practical solutions.

Presenter:

Jim DeLaHunt
Principal
Jim DeLaHunt & Associates

Track 3 - Twanguages of the World: a Language Census of Twitter
What "twanguage" do you "tweet"? Twitter, the buzzing conversation of brief web and SMS messages, exploded into wide use in 2009. But just how wide? To how many countries has it spread? And into which languages? We aimed to find out. Our "Twanguage" project is a language census on a sample of Twitter's global traffic. Come hear our findings. Which are the top languages? Are #hashtags localized? How does language correlate with location? And which Unicode character is the most rarely used? Accessible to everyone, this talk is especially interesting to students of social media and of quantative language analysis.

16:10 - 17:00   **SESSION 13**

Presenters:

Frank Yung-Fong Tang
Sr. Software Engineer
Google Inc.
Wenchao Tong
Software Engineer
Google Inc.

Track 1 - Google APIs for Text Input and Translation
In this talk, we introduce several Google public APIs to empower web developer build more powerful internationalized web site, including, but not limited to:

- Use Google AJAX Language API and element API to perform Machine Translation
- Use Google AJAX Language API and element API to empower user to input text of different language by transliteration
- Use Maps in Google Chart API and Geomap in Google Visualization API to represent information divided by geographical distribution

For each of these topics, we will first introduce the issues, following by the brief description of the API, and demonstrate with some real Google or non Google products which utilize these APIs. Short sample codes will also be walk though.

Presenter:

Roozbeh Pournader
Internationalization Specialist
HighTech Passport

Track 2 - Bidirectionalization: Demystifying Bidi Enabling
Bidirectionalization, or enabling software to be usable to people who write in bidirectional languages like Arabic and Hebrew, has sometimes been discarded as a superfluous and strenuous endeavor. This presentation will explain why bidi enabling is a must for every application and website intended for bidirectional users of the Middle East, as well as for other parts of Asia and Africa. It will also include suggestions on how to plan for, design, code, and test the bidirectionalization of such applications and sites. Last but not least, it will cover common internationalization requirements for the Middle East, including

alternative calendars, local digits, and geopolitical sensitivities.

The intended audience of this presentation are developers, software architects, and managers planning to bidirectionalize their software or add support for other requirements of the bidirectional language markets.

Presenter:

Ilya Shtein
IT Architect
Metavante

Track 3 - Banking in the Cloud: Challenges of Internationalizing Banking Software (Case Study)
Based on the experience of building the Metavante Global Banking platform, we will discuss the challenges of internationalization in a distributed, service-oriented, heterogeneous banking environment.

Internationalization in the banking industry presents a number of challenges, such as the large number of legacy applications that do not share the same terminology and the need for further terminology customization on multiple hierarchy levels, as well as transactions spanning multiple locales and time zones.

We will talk about the applicability of Unicode and Unicode standards in different architecture layers, using W3C-i18n recommendations, and discuss the effect the listed challenges have on internationalization decisions.

Program is subject to change.

- To Register for IUC33: http://www.unicodeconference.org/registration.htm
  Or, contact Suzanne Leon at suzanne@omg.org
- Exhibitor Information: http://www.unicodeconference.org/be-exhibitor.htm
  Or, contact Ken Berk at ken.berk@omg.org

- Sponsor Information: http://www.unicodeconference.org/be-sponsor.htm
  Or, Ken Berk at ken.berk@omg.org, or 781-444-0404.