Program Details

Monday, October 21, 2013

08:30-10:00	MORNING TUTORIALS
Presenter:	**Track 1: An Introduction to Writing Systems & Unicode**
Richard Ishida *Internationalization Activity Lead,* *W3C*	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:	**Track 2: Putting ICU to Work**
Steven R. Loomis *Software Engineer, IBM*	This tutorial gives attendees everything they need to know to get started with working with text in computer systems: character encoding systems, character sets, Unicode, and text processing, using the International Components for Unicode library (ICU). ICU is a very popular internationalization software solution. However, while it vastly simplifies the internationalization of products, there is a learning curve. The goal of this tutorial is to help new users of ICU install and use the library. Topics include: Installation (C++ libraries, Java .jar files, Java SPI for JDK integration), verification of installation, introduction and detailed usage analysis of ICU's frameworks (normalization, formatting, calendars, collation, transliteration). The tutorial will walk through code snippets and examples to illustrate the common usage models, followed by demonstration applications and discussion of core features and conventions, advanced techniques and how to obtain further information. It is helpful if participants are familiar with Java, C and C++ programming. Issues relating to ICU4C/C++ as well as ICU4J (Java) will be discussed. After the tutorial, participants should be able to install and use ICU for solving their internationalization problems. This updated presentation will include newer ICU features such as AlphabeticIndex and the DateTimePatternGenerator.
Presenter:	**Track 3: Internationalization: An Introduction**
Addison Phillips *Globalization Architect, Lab126 (Amazon)*	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.

10:00-10:30 - Morning Refreshments

10:30-12:00	MORNING TUTORIALS

Presenter:

Richard Ishida
Internationalization Activity Lead, W3C

Track 1: An Introduction to Writing Systems & Unicode (Cont'd.)

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:

Jim DeLaHunt
Principal, Jim DeLaHunt & Associates

Track 2: Building Multilingual Websites in Drupal 7 and Joomla 3

A practical look at the language and locale capabilities of Joomla! 3 and Drupal 7, two leading free software content management systems (CMSs). They let you build more powerful, more international websites faster. We look at: their core internationalisation and locale services, and localisation of UI and content. Each platform just had a major release, with advances in internationalisation. 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, web designers, developers, and managers of international web teams.

Presenter:
Tex Texin
Globalization Architect, Xencraft

Track 3: 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, including HTML5, 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.

12:00-13:00 - LUNCH

13:00-14:30	AFTERNOON TUTORIALS

Presenters:

Craig R. Cummings
Principal Software Engineer - Internationalization, Informatica

Michael McKenna
Globalization Engineering Leader, PayPal, Inc.

Track 1 - Unicode - The Advanced Tour

This tutorial will cover some of the more advanced subjects in Unicode. We'll discuss properties and the Unicode Code Database (UCD), normalization, character encoding forms, case mapping, boundary analysis, supplementary characters, the Unicode Collation Algorithm (UCA), Common Locale Data Repository (CLDR), and bidirectional text support, including shaping algorithms. For each, discussion will range over algorithms, implementations, pros and cons, and gotchas. We'll end with examples from real-world cases including last year's hit slide set -- 'Unicode Gone Bad'.

Presenters:

Aharon Lanin
Software Engineer, Google

Richard Ishida
Internationalization Activity Lead, W3C

Andrew Glass
Program Manger, Microsoft

Track 2 - Handling Arabic, Hebrew, and Other Right-to-Left Scripts in HTML

This tutorial explains the problems encountered when authoring web pages in a right-to-left language, or even left-to-right pages containing snippets of right-to-left text, and offers a step-by-step guide for preventing most of them. It will also highlight the relevant new features in HTML5, CSS3, and Unicode 6.3."

Presenters:

Daniel Goldschmidt
Sr. International Program Manager, Microsoft Corporation

Iris Orriss
Internationalization Manager,

Track 3 - Localization Workshop

Two highly experienced industry experts will illuminate the basics of localization for session participants over the course of three one-hour blocks. This instruction is particularly oriented to participants who are new to localization. Participants will gain a broad overview of the localization task set, issues and tools. Subjects covered will be fundamental problems that localization addresses such as components of localization projects, localization tools and localization project management. There will also be time for questions and answers plus the opportunity to take individual questions offline

Facebook

with the presenters.

14:30-15:00 - Afternoon Refreshments

15:00-16:30 **AFTERNOON TUTORIALS**

Presenter:

Martin J. Dürst
Professor, Aoyama Gakuin University

Track 1 - Character Normalization

This part of the tutorial covers character normalization in Unicode. It starts with a short history of normalization to help participants understand the purpose of normalization. We will then introduce the four standard normalization forms (NFC, NFCK, NFD, NFDK), followed by (non-official!) normalization variants and other normalization operations such as quick checks. Next, we cover normalization pitfalls such as the interaction of normalization with other operations (e.g. string concatenation, casing) and Unicode versions and corrigenda. We continue by discussing implementation-related details including where to find the necessary data for implementation and testing, and various trade-offs for efficient implementations. We conclude giving guidelines and advice on where to use which form of normalization (and where to not use normalization).

Presenter:

Norbert Lindenberg
Internationalization Consultant, Lindenberg Software LLC

Track 2 - Internationalizing JavaScript Applications

Somewhere on the way to global success, you'll have to get your software ready to support different human languages and cultures. For software written in JavaScript it's not obvious how to start: While the new ECMAScript Internationalization API Specification defines core functionality, it's not available in all browsers yet, and doesn't cover all the functionality you need. This talk surveys what browsers and libraries provide today to make the task easier, and how you can best take advantage of them.

Topics include resource loading, message construction, sorting, number formatting and date and time formatting, support for emoji and other supplementary characters, regular expressions, Unicode normalization, and internationalized domain names.

Presenter:

Daniel Goldschmidt
Sr. International Program Manager, Microsoft Corporation

Iris Orriss
Internationalization Manager, Facebook

Track 3 - Localization Workshop (Cont.)

Two highly experienced industry experts will illuminate the basics of localization for session participants over the course of three one-hour blocks. This instruction is particularly oriented to participants who are new to localization. Participants will gain a broad overview of the localization task set, issues and tools. Subjects covered will be fundamental problems that localization addresses such as components of localization projects, localization tools and localization project management. There will also be time for questions and answers plus the opportunity to take individual questions offline with the presenters.

Tuesday, October 22, 2013

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

09:15-10:00 ***KEYNOTE PRESENTATION - Enabling the Next Billion Multilingual Users for Wikipedia***

Presenter:

Alolita Sharma

Director of Engineering, Wikipedia

Wikipedia is one of the cornerstones of the Web today. People like you and I contribute an amazing variety of interesting content in the form of articles, photos and quotations in hundreds of languages across the globe.

Wikipedia is building ground breaking tools and technologies for hundreds of languages. These open source tools enable billions of users to read Wikipedia articles in their own language with web fonts available on-demand. These also enable contributors to edit Wikipedia articles using input methods and onscreen keymaps on desktop and mobile platforms. Other innovative language tools we're building will help editors collaborate and translate article content.

10:00-10:30 - Morning Refreshments

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

Presenter:

Roy Yokoyama
Principal Global Engineer, Google-Motorola Mobility

Track 1 - Internationalization and Localization for Android Devices

Android is the world's most popular mobile platform. There are more than 600,000 apps and games available worldwide and growing strong. This panel will cover the anatomy of Android localization frameworks, structure of resource assets and qualifiers, and how to localize your applications.

Presenter:

Peter Edberg
Senior Software Engineer, Apple Inc.

Track 2 - What's New in CLDR

The Unicode Consortium's Common Locale Data Repository project (CLDR) defines LDML (Locale Data Markup Language) and uses it to organize and provide the most extensive open repository of locale data, with data collected primarily via the web-based Survey Tool. This session provides a brief

Mark Davis
Sr. Internationalization Architect, Google Inc.

Steven R. Loomis
Software Engineer, IBM

overview of CLDR, then focuses on recent and forthcoming enhancements, including more consistent non-Gregorian calendar support and addition of the Korean Dangi lunar calendar, formatting time durations, collation data changes, more control of currency formatting, improvements to the data collection process, and JSON support. A significant amount of time will be reserved for demos of behavior based on CLDR data.

Presenters:

Addison Phillips
Globalization Architect, Lab126 (Amazon)

Richard Ishida
Internationalization Activity Lead, W3C

Track 3 - The Multilingual Web: A Status Report

New standards activity at the W3C and other standards bodies is helping improve the globalization of the Web. From HTML5 to CSS3, from JavaScript to Unicode bidi, support for creating international content and Web-based applications is evolving quickly. This presentation, from W3C internationalization leaders Addison Phillips and Richard Ishida explores the changes that are available today, the status of on-going work, and the challenges that remain.

11:30-12:20	SESSION 2

Presenter:

Michael S. Kaplan
Program Manager, Microsoft

Track 1 - The Story of MSKLC

The Microsoft Keyboard Layout Creator (MSKLC) been around since 2004, and has been downloaded over two million times. This talk is going to be about where MSKLC came from and where it is going to over the next year and beyond. Find out before anyone else by being here for *this* presentation...

Presenters:

Peter Edberg
Senior Software Engineer, Apple Inc.

Mark Davis
Sr. Internationalization Architect, Google Inc.

Steven R. Loomis
Software Engineer, IBM

Cameron Dutro
Senior Software Engineer, International Engineering Team, Twitter

Track 2 - CLDR Users Panel and Discussion
Coming Soon...

Presenter:

Felix Sasaki
DFKI / W3C Fellow

Christian Lieske
SAP AG

Track 3 - ITS 2.0: Facilitating Automated Creation and Processing of Multilingual Web Content

The Internationalization Tag Set (ITS) 2.0 http://www.w3.org/TR/its20/ is an upcoming technology developed by the World Wide Web Consortium (W3C). Its predecessor, ITS 1.0, provides a set of metadata items ("data categories") for creating internationalized XML which can be localized effectively. ITS 2.0 extends the scope of ITS 1.0 in various ways. First, ITS 2.0 can be applied to further formats, e.g. HTML5 or other versions of HTML. Second, ITS 2.0 provides additional data categories (one aim is to create further application areas for ITS in the realm of automated language processing such as machine translation, text analysis annotation or workflows including automatic language quality checks). Third, ITS 2.0 provides implementations in interoperability usage scenarios. A dedicated draft document http://www.w3.org/TR/mlw-metadata-us-impl/ already has been published to provide information about these scenarios. This presentation will 1) introduce the basic principles of ITS 2.0; 2) describe its relation to technologies like HTML5 and XLIFF; 3) exemplify selected data categories and usage scenarios; and 4) discuss what ITS 2.0 cannot achieve and what might be desirable in the future.

12:30-13:30 - LUNCH

13:30-14:20	SESSION 3

Presenter:

Jeremy Cole
Sr. System Engineer, Google, Inc.

Track 1 - MySQL History and Practical i18n and i10n

MySQL is one of the world's most popular relational database systems, powered many of the world's busiest websites. As the world has moved toward Unicode,MySQL has had to do so as well. Learn about the challenges specific to MySQL, general database challenges, solutions, and pitfalls for internationalization of databases and internationalization and localization of MySQL itself. MySQL has not always made the best decisions in its support of Unicode, so we'll also look at some of the snags MySQL has encountered along the way.

Presenter:

John Emmons

Track 2 - CLDR Data for Javascript and the Web

The Unicode Common Locale Data Repository (CLDR) is the most extensive and widely used standard

Senior Software Engineer, IBM

repository of locale data for software developers to use as a resource in building internationalized applications. The data is typically published in XML format as defined by the Locale Data Markup Language (LDML) specification (Unicode Technical Report TR#35). XML format, however, is not always the best choice when it comes to writing applications for the web. This presentation will highlight recent efforts by the CLDR technical committee to create and maintain an algorithm and tool set to convert CLDR's data to a standardized JSON format that would be consumable by Javascript applications, while still being able to represent all the various types of data that are currently maintained in the CLDR.

Presenter:

Track 3 - Unicode Programming in Modern Perl

Nick Patch
Software Engineer, International, Shutterstock

In 2010 the Perl 5 language switched to a yearly major release cycle with monthly developer releases. Perl has a history of great Unicode support and the new development process has enabled the language to rapidly enhance Unicode functionality and support new Unicode standards. This talk will demonstrate the current state of Unicode in Perl, review the exciting changes in the last few years, and touch on future development.

The following core language features will be discussed:

- character strings
- character properties
- regular expressions
- case mapping and case folding
- normalization
- extended grapheme clusters
- segmentation
- identifier names
- UTF-8 source code
- I/O encoding layers

14:30-15:20 SESSION 4

Presenter:

Track 1 - No More Tofus and Ransom Note: Beautiful Fonts for All

Jungshik Shin
Systems Engineer, Google, Inc.

Unicode has over 110 thousand characters in over 100 scripts for all the languages. No set of fonts covers them all with the consistent and harmonious look and feel. Instead, we have a lot of fragmented fonts with a hodgepodge coverage with a lot of "holes" leading to Tofus and "ransom note-like" rendering. Google embarked on a very challenging endeavor of developing a set of fonts to cover all the characters in the Unicode with beauty and harmony across scripts. In this talk, we will present what we have accomplished and how we have overcome multitude of challenges and difficulties, expected and unexpected, in our journey to "Beautiful Fonts for All".

Presenter:

Track 2 - Do You Speak CDLR?

George Rhoten
Language Technologies Engineer, Apple Inc.

CLDR (Common Locale Data Repository) is a wonderful resource for localized data. It has a vast supply of localized data for dates, times, numbers, names and related data. Most of this data is geared towards the printed world of the Internet and visual applications.

The printed format of this data is typically short hand notation for concepts that are speakable. How would you say a number like the number 1? How do you say the time like 12:00 PM? What is the pronunciation of a date like January 1? What is the pronunciation of a currency value like $1.99? This presentation covers pronunciation topics that are being planned for CLDR.

Presenter:

Track 3 - Implementing Normalization in Pure Ruby - the Fast and Easy Way

Martin J. Dürst
Professor, Aoyama Gakuin University

Unicode normalization eliminates encoding variants for what is essentially the same character. We present a new efficient way to implement normalization in scripting languages such as Ruby. Our implementation takes advantage of two facts: First, that large percentages of many kinds of text do not need to be changed when normalizing, and second, that the number of different sequences that need to be normalized in a certain language or group of languages is always much lower than the number of theoretically possible sequences. We exploit each of these facts with built-in functionality of Ruby. Specially-crafted regular expressions capture sequences of characters in possible need for normalization, and a hash structure is used to look up their normalization. The hash structure starts empty and is updated by a default block of code whenever a new sequence of characters needing normalization is identified. As a result, no actual normalization code is executed once the hash structure incorporates the character sequences used in a particular language.

Performance tests with many different implementations and a wide range of languages (German, Vietnamese, Korean,...) with different normalization needs show that our method is one or more orders of magnitude faster than straightforward implementations in scripting languages, and in some cases close to the speed of implementations in compiled languages. Because the code is written in pure Ruby, it is very easily portable and configurable, and similars techniques may be useful for other internationalization operations such as case conversions.

15:20-16:00 - Afternoon Refreshments

SESSION 5

Presenter:

Vladimir Weinstein
Engineering Manager, Google Inc.

Mark Davis
Sr. Internationalization Architect, Google Inc.

Track 1 - Innovations in Internationalization at Google

This presentation covers the range of internationalization challenges that Google has encountered and overcome in the past year, illustrating the challenges that many companies face.

Among the topics are how to better serve multilingual users (personal names, plurals, gender, language resolution), localization of entities (http://goo.gl/3s7SR), phone numbers and addresses, expanding CLDR and ICU (and Google products) to more languages, speech-to-text, machine translation, input & fonts, client-side i18n, and the overall localization process.

Presenter:

Shan Fu
International QE, Adobe Systems

Jing Lai
International QE Team Lead, Adobe Systems

Track 2 - Best Practice in A New Working Model for Localization Testing

Localization testing of multilingual software is important for a company's international marketing strategy. Nowadays, language-oriented working model is widely used in software localization functional testing; however, this model doesn't work as expected in testing efficiency and effectiveness in practice. This proposal introduces a new working model for localization functional testing to improve the product quality and reduce the testing cost by taking one Adobe product localization functional testing as an example; what's more, it also discusses how to choose a proper working model for large-scale localization functional testing.

Unlike the original "language-oriented" model, this new working model is based on feature-oriented model, and has been improved according to team size, languages, product type, testing phase and other factors. This enhanced working model has proved to be effective in localization functional testing on an Adobe enterprise desktop product, which supports over 20 languages.

As an enhanced feature-oriented working model, its main advantages are as follows:

1. Better product quality. With this model, quality engineers could have more opportunities to execute new feature test cases, so they could find more bugs on localization, even on core functions.

2. Higher efficiency, lower cost. In main release of the Adobe product, around 50% of features are new ones, so the cost of time for new features learning is very high. In terms of localization functional testing on new features, the enhanced feature-oriented working model could help reduce working hours by around 30% as opposed to language-oriented one.

3. More helpful for team building. With this model, quality engineers could have a chance to study product features and discuss testing issues in depth, so that both individual capability and team building could be improved.

Practice shows that with this new working model, 15% of cost on localization functional testing has been reduced, and 40% more bugs have been found, which turns out to be more effective than before; besides, it also has a huge business value to a software product targeting international markets.

Presenters:

Tex Texin
Globalization Architect, Xencraft

Craig R. Cummings
Principal Software Engineer - Internationalization, Informatica

Track 3 - Comparing Javascript Libraries

Which JavaScript library is best for international deployment? This session presents the results of an investigation of the features of several JavaScript libraries and their suitability for international markets. We will show how the libraries were tested and compare the results for at least: Dojo, jQuery, and YUI. The results may surprise you and will be useful to anyone designing new international or multilingual JavaScript applications or supporting existing ones.

17:00-17:50 **SESSION 6**

Presenter:

Behnam Esfahbod
Research Assistant, Xencraft

Track 1 - Arabic Script SHOULD NOT be so Scary!

Arabic script has faced radical changes in the past century that have left it in a critical situation in the digital era. Writing styles for the script look fairly different from region to region, yet it always is considered the same script in the mind of its natives. Logical encoding of the script and existence of similar characters in Unicode Arabic blocks have caused serious challenges for most of multilingual applications.

Providing a model for the textual part of human-computer interaction, we model security and usability issues of Latin and Arabic scripts, demonstrate the similarity of the problems in these two scripts, and introduce the Arabic Script Shape Mapping algorithm as a unified method for applications - like internationalized domain names and user identifiers - to handle such problems. The algorithm is designed for simplicity and behaves similar to how Unicode Case Mapping algorithm works for Latin script and the like.

Shifting the solution from the servers to applications, we show that how this algorithm can be used to maintain enhance usability for average users, whilst eliminate the need for complicated methods like "bundling" used in internationalized domain names.

Presenter:

Tex Texin
Globalization Architect, Xencraft

Track 2 - Critical Values for i18n Testing

This presentation is an extension of the popular presentation from last year "Does it Hurt When I do This? Data for I18n Testing".

In this session, we recommend specific data values that are likely to identify internationalization problems in software intended for global markets.

Based on years of global software experience, these data values are useful for functional or linguistic QA tests of internationalized software. In the previous session, data value recommendations included character encoding, postal address, locale and other data types typically used in software and will trigger common internationalization problems. This presentation will offer specific new test suggestions.

Presenters:

Norbert Lindenberg
Internationalization Consultant, Lindenberg Software LLC

Nebojša Ćirić
Software Engineer, Google, Inc.

Zbigniew Braniecki
Software Engineer, Mozilla Corporation

Track 3A - Internationalizing the Core of JavaScript, Second Edition

After last year's approval of the first edition of the ECMAScript Internationalization API Specification, internationalization work on the core of JavaScript continued in three areas: Implementation of the API in browsers, broadening its scope in a second edition of the specification, and strengthening Unicode support in the ECMAScript Language Specification. We provide an update on the latest work in all three areas.

Track 3B - L20n, Next Generation Localization Framework for the Web

As the web grows and becomes more dynamic, Mozilla seeks to uphold the standards and practices that keep it open for all.

The key to that is to break the paradigm that linguistically ties translations to a source string. That paradigm, which is fundamental to almost all localization technologies used on the Web today, is becoming more limiting than ever, especially with the rise of responsive user interfaces.

Based on 15 years of experience in building open source software, localized into over 90 locales, Mozilla came up with a new localization framework that shifts this paradigm, called L20n. L20n isolates localizations and enables translators to provide naturally expressive translations for even the most complex user interfaces.

Mozilla is investing in moving its products - Firefox, Firefox OS, and Firefox for Android - to this new architecture. At this talk you'll get a chance to see what L20n brings to the table.

18:00-19:00 - IUC37 CONFERENCE RECEPTION

Wednesday, October 23, 2013

09:00-09:50 | **SESSION 7**

Presenter:

Adam Asnes
President, Lingoport

Track 1 - A Cost and ROI Model of Internationalization Issues

In many organizations, localization engineering teams and i18n engineers must battle to gain developer attention and organizational focus. Often the i18n Lead, if there is one, does not independently command budget initiatives. Yet i18n and localization engineering issues that are found during testing or after the "English" release are tolerated and even considered normal process. When compared to better-funded development initiatives like security, we've heard statements like "nobody gets into the news because of an encoding issue."

The costs in engineering and time are not well understood, and we think if they are clearly accounted for and calculated in a flexible system, we can further the case for software globalization planning and efficiencies at the highest management levels.

Working in association with customers at Intel, the LDS Church and others, we have catalogued the most pervasive types of globalization engineering issues and their effects in an engineering ROI calculator. The calculator accounts for broad scope, variable engineering costs, the time during the development process issues are found and fixed, cost in time and money, and information about delays in planned revenues.

We would like to share this calculator model with the industry to help conference participants and their organizations understand and communicate financial consequences in time, money and strategic objectives.

Presenter:

Travis Keep
Software Engineer, Google, Inc.

Peter Edberg
Senior Software Engineer, Apple Inc.

Yoshito Umaoka
Senior Software Engineer, IBM

Track 2 - A Time for Everything

There is a time for everything. Dates and times are an important part of our everyday lives, and displaying them correctly in software is essential. This session will cover some of the subtle details associated with formatting dates and times as well as how the ICU API abstracts away many of these details while still remaining powerful and flexible. Topics include adapting localized formats to a client specification of which fields to include, formatting intervals and durations, handling non-Gregorian calendars, handling special number systems in date formats, and multiple ways of formatting timezone names.

Presenters:

Deborah Anderson
Researcher
University of California Berkeley

Anshuman Pandey
Doctoral Student, Dept. of History, University of Michigan

Track 3 - The Script Encoding Initiative: The Successes and Tribulations of Encoding Scripts

Since 2002, the UC Berkeley Script Encoding Initiative has been assisting user communities to get various eligible characters and scripts proposed for inclusion in the Unicode Standard. This joint presentation, given by the Project Leader of SEI and one of its key Unicode proposal authors, will discuss the highlights as well as the outstanding problems faced by those proposing unencoded scripts.

The SEI project has been very successful, with over 69 scripts approved by the Unicode Technical Committee. The project has also sponsored work on over 38 preliminary script proposals. (These proposals require additional information and/or user community buy-in before they are considered mature and ready for approval by the standards committees.)

The project's work on final and preliminary proposals has been made possible thanks to funding from NEH, Google, and individual and group donors. However, the NEH and Google support has come to an end, so funding is limited. This presents a problem, for script research and the standards approval process itself take several years. Long-term support is the key to ensuring eligible scripts are kept in the encoding pipeline. The SEI Project Leader will suggest a few ideas on possible approaches to acquiring stable funding.

After 11 years, it would seem the list of unencoded scripts should be significantly reduced, but because of the development of new scripts and increased interest (and recognition of the importance of encoding) of historic scripts, the number of unencoded scripts is at about 90. The co-presenter, Anshuman Pandey, will discuss the creation and propagation of newly constructed scripts, a phenomenon found particularly in South Asia. He will also discuss problems in getting newly approved scripts supported in commercial software and fonts, which hobbles the user communities.

The talk will end with some suggestions on how members of the audience can assist in the effort to get unencoded scripts into Unicode, and be supported in software.

10:00-10:50 SESSION 8

Presenter:

Loïc Dufresne de Virel
Localization Strategist, Intel Corporation

Track 1 - From User Experience to Benutzererlebnis - UX Joins L10N to Offer a Better Global User Experience

There is much to gain by combining UX and L10N, often afterthoughts in the world of software development, to provide a better global user experience. From basic internationalization to customization and culturalization, we'll cover topics that seem trivial but present some significant challenges from a User Experience and Localization standpoint, such as organization of information, methods of payment, collection of personal information, handling of genders, numbers, and declensions, or language selection. We'll then look into the new usage models offered by technologies such as Voice Recognition or Perceptual Computing, which promise to change the way we interact with our devices.

Presenters:

Addison Phillips
Globalization Architect, Lab126 (Amazon)

Track 2 - How to Spot a Flying Unicorn: Analyzing User Interface Designs for Localizability

The user interface is often the difference between success and irrelevance for a software product. Getting the maximum of usability, accessibility, and functionality out of design, with a minimum of user befuddlement is a big job.

But what happens when that user interface will be translated into 20 languages? Will the design accommodate their different needs? Will they even fit onto the screen? Are cultural or linguistic issues even addressable at design time? How can the designers, developers, or the localizers see these "flying unicorns" lurking in their design choices without actually building the different language products?

This presentation gives concrete examples from real products showing an approach to reviewing user experience design for a global audience.

Presenters:

Brian Kemler
Program Manager, Google, Inc.

Track 3 - Burmese - Challenges & Lessons Learned from an Early Adopter

Google Search is the first major online service to support the Burmese language in its interface. Although the Unicode standard includes a block for the Burmese writing system, Burmese text in

Craig Cornelius
i18n Engineer, Google, Inc.

websites is written using multiple, non-unicode schemes including Zawgyi and other font encodings. Enthusiastic speakers developed fonts in the vacuum left by major vendors due to fear over trade sanctions or lack of a perceived market importance. Now that Myanmar is opening, interesting challenges arise.

Detecting the encoding methods and processing Burmese text present some intriguing challenges around segmentation, input and output methods. Our discusses these themes, and outlines our progress toward encouraging the standardization of the Burmese language on the web. Of particular importance is promoting the adoption of the Unicode font MM3 within the context of not only Burmese, but also minority languages in Myanmar such as; Karen, Kayah, Pali, Rumai Palaung, and Shan.

10:50-11:10 - Morning Refreshments

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

Presenter:

John Huan Vu
Senior Software Engineer, Zynga Inc.

Track 1 - Tech View of Global Social Games Development

This session will cover many of the technology problems and many of the solutions involved in creating, managing, and deploying social games for the World. Zynga creates games in Flash for Facebook and the Web, and mobile games for iOS and Android using a range of different technologies. But globalization, to be effective in this environment, must follow a consistent model that is portable, flexible, and meets the needs of the many issues games encouter when designed for multiple languages and cultures. We will cover the following topics:

- Design issues for global content and global asset database design
- Linguistic issues for social dialog across langauges
- What to track for effective metrics
- Accelerated development life cycles
- Cadence and content release driving blinding speed of process
- Build systems and process automation
- Localization test automation (where possible)
- Portability issues for core technology and build systems across development environments
- Continuous improvement

After this talk, you should have a pretty good idea of what its like to be in the midst of international engineering and production at a fast (really fast) paced social gaming company, publishing dailiy in multiple languages and cultures.

Presenters:
Michael Kuperstein
Localization Engineer, Intel

Octavio Ramos
Localization Software Quality Assurance Lead & Engineer, Intel

Track 2 - Which *Bleep* Language Tag Should I Use?

How hard could it possibly be to choose a list of language abbreviations for your product? It seems pretty easy, since most of the new development environments all claim to support BCP 47 / RFC 5646. But as it turns out, language tag choice is much harder than it seems, since there isn't a definitive list of tags in the BCP 47 specification. Therefore, every company has their own list of language tags, and some are very, very creative! Besides, the end user will probably never see these language tags, so it's not that important, right?

Unfortunately, choosing the wrong language tag can cause technical failures in products, or can cost many thousands of dollars in lost translation memory leveraging, or can create confusion in teams that generates hundreds of emails and awkward workarounds. Perhaps you have installed the Chinese Traditional language pack for Windows, and your application is displaying English instead? Maybe you have localized into French for Canada, and the app refuses to show the translations? Customers may ask, "We want Spanish for Latin America!" Ok then, what is the language tag for Spanish in Latin America - for each of the dozen systems that process language data in your company? What do you do?!?

In this entertaining session, we will place giant red construction cones around the potholes of choosing the wrong language tags when developing drivers, desktop apps, and web applications on various operating systems, or when transferring data between different enterprise and operating systems.

Presenter:

Thomas Milo
Partner, DecoType

Track 3 - **The Qur'ān Project**

This is a field report about a project initiated by the Sultanate of Oman. The project aims to display, search and quote Qur'ān text on the web in a typographically stable and orthographically flawless manner, regardless the operating system or the type of web device. After a general description of the project, two areas will be discussed in more detail:

- Scalable Vector Graphics (SVG) technology will be adapted to stabilize the typography. Browser deficiencies regarding SVG will be identified; SVG issues with Unicode will be identified.

- Typography issues regarding Arabic orthography in general and Qur'ān in particular will be identified, particularly the amphibious letters, not handled by any software to date.

12:00-13:00 - LUNCH

SESSION 10

Presenters:

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

Track 1 - What's New in 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, from mobile phones like Android or iPhone all the way up to mainframes and cloud server farms. 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 recent updates in ICU 51 & 52, including the latest support for Unicode 6.3 and CLDR 24, date/time formatting & parsing improvements, and other changes (see http://site.icu-project.org/download). The presentation will also touch on ICU's planned direction for future releases.

Presenter:

Joel Sahleen
Globalization Engineer, Adobe Systems

Track 2 - A Web With Many Voices: The Theory and Practice of Multilingual Hypertext

This presentation deals with the theory and practice of multilingual hypertext. The goal of the presentation is to develop a rudimentary theory of "multilingual hypertextuality" and then use this theory to analyze certain practical issues in the areas of hypertext internationalization and localization. I will start by reviewing Jorge J.E. Gracia's book A Theory of Textuality and Sergio Cicconi's article "Hypertextuality" so that I can explain what I mean by these odd-sounding terms and show how they are used in contemporary philosophy and critical theory. Having laid the theoretical groundwork, I will then introduce the concept of "multilingualism" and discuss what is involved in delivering hypertextual content that can be dynamically rendered in different languages. In the area of text internationalization, I will focus on the issues of externalization, composition, identification, organization and annotation. In the area of text localization, I will focus on the issues of text routing, retrieval and rendering. The presentation will conclude by briefly discussing how to integrate multilingual hypertext in different types of application architectures.

Presenters:

Markus Scherer
Unicode Software Engineer, Google, Inc.

Mark Davis
Sr. Internationalization Architect, Google Inc.

Track 3 - One Hundred Ways of Sorting Strings (and Counting)

How do we bring order to 100,000 characters? How do we support 100 language-specific sort orders with little data, and make it fast? How do we build phone book indexes, put your native script first, ignore accents and punctuation, and search for words in a web page?

Join us for a tour of the Unicode Collation Algorithm (UCA), the CLDR collation data, and the ICU APIs for sorting, indexing and searching.

14:00-14:50 SESSION 11

Presenter:

Roshan Singh
MTS 2, Adobe Systems

Track 1 - NLP: A Quick Global Multilingual Approach to Develop Named Entity Recognizer (NER) Using DBpedia

Named-entity recognition (NER) (also known as entity identification and entity extraction) is a subtask of information extraction that seeks to locate and classify atomic elements in text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. It forms the core of Semantic Text analytics applications. NER systems have been created that use linguistic grammar-based techniques as well as statistical models. Hand-crafted grammar-based systems typically obtain better precision, but at the cost of lower recall and months of work by experienced computational linguists. Statistical NER systems typically require a large amount of manually annotated training data. We discuss DBpedia's multilingual data resource and techniques to extract meaningful information to develop NER which can be extended to all major languages like German, Chinese, Japanese, etc. apart from English. And hence we share a quick global approach to develop a simple NER out of DBpedia with least effort for all major languages (EN, DE, ZH, JA, etc.) which tags more than 10 types of named entities like Person, Place, Organization, Work, Medicine, Transport (vehicle), etc. We demonstrate the NER for EN and DE on various test articles.

Presenter:

John Huan Vu
Senior Software Engineer, Zynga Inc.

Track 2 - Localization Framework for Dynamic Text

"Localization Framework for Dynamic Text" is a pending patent that was developed on the Mafia Wars team at Zynga. During the time, Mafia Wars could not scale to a typical localization framework due to the fact that (1) the game runs mostly on PHP with no mature internationalization framework at the time, (2) the game generates a lot of dynamic text with at least 4.5 million words, and (3) the game's codebase is over 2 years old. This innovative framework addresses the problems by (1) capturing the most up-to-date generated dynamic strings, (2) scaling to an existing (legacy) codebase, (3) supporting engineers with little or no localization knowledge, and (4) enabling a continuous localization process for new features and content created. In this talk, you will learn about this innovative framework in detail and how it can be utilized as an alternate to a typical localization framework.

Presenter:

Addison Phillips
Globalization Architect, Lab126 (Amazon)

Track 3 - A Dog's Breakfast: Sorting Earth's Largest Collection of Content Without Going Crazy

Sorting is one of the most basic, fundamental things we do with text. When you have lots of data, putting it into some kind of order makes it easier for people to find, access, and use it.

Your Amazon Kindle can store hundreds of books (videos, MP3s, etc.), in many languages used around the globe. How would you sort "Earth's Biggest Selection"? How can you sort many languages of content at the same time?

The Unicode Collation Algorithm is a start. But much more is required. This session gives the basic "tutorial facts" about sorting, and then delves into the additional issues that surround handling user content--and how we turned a real mess (a "dog's breakfast") into something users could navigate.

14:50 – 15:10 - Afternoon Refreshments

15:10 - 16:00	SESSION 12

Presenter:
Roshan Singh
MTS 2, Adobe Systems

Track 1 - Text Analytics: Multilingual Dynamic Ontology Creation Using Wikipedia

Ontology provides controlled, consistent vocabularies to describe concepts and relationships. It forms the core of Semantic Text analytics applications. Currently building ontology is complex, manual and time consuming. We discuss Wikipedia's multilingual data resource structured in the form of complex directed graph and issues in parsing the graph to extract meaningful information and ontology tree which can be extended to all major languages like German, Chinese, Japanese, etc. apart from English. We have developed ontology in EN, DE, ZH and JA and the solution can be extended to all other major languages supported by wikipedia.

We demonstrate "Manthan" (the document categorizer), an application of this dynamic Ontology, to classify/categorize arbitrary documents in EN and DE. The result section also visually displays the tree hierarchy of the category path extracted out of ontology.

Presenter:

Alolita Sharma
Director of Engineering, Wikipedia

Track 2 - Making Agile Work for Global Software Development i18n Teams @Wikipedia

Innovation is key to software engineering success. Wikipedia's innovative language engineering projects use agile development to create high-quality internationalization and localization features and apps in faster iterations. This talk examines what works and what doesn't when using agile development for large open source projects with globally distributed teams. This talk will help developers and engineering managers better implement a successful agile process for their open source projects especially for internationalization and localization engineering.

Presenter:

Andrew Glass
Program Manger, Microsoft

Track 3 - International Features in Windows 8.1

For the past fifteen years Windows has expanded its language support with each release. Core areas of improvement have been text display, locale data, and input. The upcoming release of Windows 8.1 continues this tradition by adding to the number of supported locales, keyboards, and writing systems. Highlights of international features in Windows 8.1 include new support for Javanese and Buginese. Another noteworthy improvement is new support to render the Syriac Abbreviation Mark via OpenType. In addition to these, there have been numerous incremental updates to other writing systems for Unicode 6.3. This presentation will discuss these and other international components new in Windows 8.1.

16:10 - 17:00	SESSION 13

Presenter:

Claudia Galván
Technical Advisor

Track 1 - Going Global from the Ground Up!

Imagine you have the opportunity to build the international plan and implementing it from the ground up! In this talk I share my latest internationalization project in a late stage startup company in Silicon Valley. From building the market roadmap, legal, finance, engineering, QA, operations and even search and PR with no dedicated international team. Having led international teams at Oracle, Adobe and Microsoft, I leveraged many of the industry best practices and even developed new ones.

Presenter:

Gaurav Bathla
Quality Engineer Lead Globalization, Adobe Systems

Track 2 - Overcoming Testing Challenges with Localization in Agile Model for Adobe Products

This topic will cover various challenges in the localization cycle of products using agile/ scrum workflow and their mitigation. While most of the products localized today use the agile-scrum model; the challenges include optimizing the localization testing process w.r.t. core development cycle, re-work effort, bugs deferred due to time constraint, and the most feared last minute changes amongst others. After studying the bugs/ issues encountered over various localization cycles; we ran a pilot project where the approach was changed to using a modified variant of the V-model with the scrum model in order to reach a resolution to counter the various concerns.

With localization of various Adobe products reaching millions of customers; nearly 26 locales across multiple Windows and Mac platforms get modified for any change in the English version of the product; increasing the complexity as the change needs to be tested across a matrix of around 300 variants. The situation moves towards a nail biting end; around the release time frame of the product. With the agile-scrum model these changes come regularly and with a shorter turnaround time; thus generating a need for a better managed model for localization.

Reviewing the challenges faced and the adhoc solutions employed over the past cycles for various Adobe products; a strategic change to the approach of the agile development process was tested. As part of the change every module of the product was treated as a sovereign artifact complete with

exclusive development, testing, UI freeze and release cycles before integrating the same into the product mainline independent of the scrum timelines. These modules were strictly frozen from development perspective once integrated. This methodology along with multiple other process enhancements reduced the re-work timeframe to a miniscule level as compared to earlier cycle in addition to a huge saving in localization efforts.

Presenter:

David Lemon
Senior Manager, Type Development, Adobe Systems

Track 3 - Opening New Doors for Fonts

The "PostScript" (CFF) font format, in which most of the world's fonts are developed, is commonly used for all the traditional forms of graphic design, such as books, magazines, newspapers, advertising, posters, logos, packaging, and movie titling. But for the most part it hasn't been used in HTML pages or on mobile devices. Those environments have often done a poor job of displaying the fonts in this format, so designers have been limited to using only TrueType. Because TrueType is harder to develop and produces larger fonts, there are advantages to being able to use CFF as well. Adobe and Google have been working with the developers of FreeType, the open-source font rendering engine used in billions of devices, to improve the font imaging solutions available to browsers and mobile devices. David Lemon will talk about the improvements coming soon to a screen near you, what this means for designers and developers, and also discuss how companies can work together to bring value to type users via open-source offerings.

Program is subject to change.

- To Register for IUC37: http://www.unicodeconference.org/registration.htm
  Or, contact Maureen Kaizer maureen@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, contact Ken Berk at ken.berk@omg.org, or 781-444-0404.