Program Details

Monday, November 3, 2014

08:30-10:00	MORNING TUTORIALS

Track 1: An Introduction to Writing Systems & Unicode

Presenter:

Richard Ishida
Internationalization Activity Lead, W3C

This 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 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.

Track 2: Localization Workshop

Presenters:

Daniel Goldschmidt
Sr. International Program Manager, Microsoft Corporation

Iris Orriss
Internationalization Manager, Facebook

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.

Track 3: Web Internationalization - Standards and Best Practices

Presenter:

Tex Texin
Globalization Architect, Xencraft

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

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

Presenter:

Richard Ishida
Internationalization Activity Lead, W3C

This 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 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.

Presenters:

Daniel Goldschmidt
Sr. International Program Manager, Microsoft Corporation

Iris Orriss
Internationalization Manager, Facebook

Track 2: 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.

Presenter:

Jim DeLaHunt
Principal, Jim DeLaHunt & Associates

Track 3: 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 internationalization and locale services, and localization of UI and content. Each platform just had a major release, with advances in internationalization. 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 practitioners. A good tutorial for web site product managers, web designers, developers, and managers of international web teams.

12:00-13:00 - LUNCH

13:00-14:30	**AFTERNOON TUTORIALS**

Track 1 - Unicode - The Advanced Tour

Presenters:

Craig R. Cummings
Principal Software Engineer - Internationalization, Informatica

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 the popular hit slide set -- 'Unicode Gone Bad'.

Michael McKenna
Globalization Engineering Leader, PayPal, Inc.

Presenter:

Addison Phillips
Globalization Architect, Lab126 (Amazon)

Track 2 - Internationalization: An Introduction

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.

Presenters:

Steven R. Loomis
Software Engineer, IBM

Marco Fossati
Researcher, SpazioDati

Sebastian Hellmann
Researcher, AKSW Research Group

Track 3 - What You Can Make Out of Linked Data - General Introduction and the DBpedia+ULI Use Case

The aim of this tutorial is threefold. First, we will introduce linked open data: what is it about & used for, what are technical building blocks and common use cases?
Then we will dive into a key data set in the linked open data cloud: DBpedia. DBpedia is a crowd-sourced community effort to extract linked data from Wikipedia and make this information available on the Web. We will introduce DBpedia as a data resource, as a tooling infrastructure, and as a community. The community has been formalized as the DBpedia Association and is represented as of writing by 14 countries (still growing).

Finally we will dive into the specific use case for DBpedia: its application within the Unicode Localization Interoperability Technical Committee (ULI). ULI uses DBpedia's structured data to build an abbreviation data base. That data base then supports sentence boundary detection processes. This is just a seed use case for DBpedia in the context of ULI, with potential applications in many other areas.

14:30-15:00 - Afternoon Refreshments

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

Track 1 - Tailoring Collation to Users and Languages

Presenter:

Markus Scherer
Unicode Software Engineer, Google, Inc.

This interactive session shows how to use Unicode and CLDR collation algorithms and data for multilingual sorting and searching. Parametric collation settings - "ignore punctuation", "uppercase first" and others - are explained and their effects demonstrated. Then we discuss language-specific sort orders and search comparison mappings, why we need them, how to determine what to change, and how to write CLDR tailoring rules for them.

We will examine charts and data files, and experiment with online demos. On request, we can discuss implementation techniques at a high level, but no source code shall be harmed during this session.

Presenters:

Jan Anders Nelson
Senior Program Manager, Microsoft

Cameron Lerum
Senior Program Manager, Microsoft

Track 2 - Learn to Easily Create Windows Desktop, Store and Phone Multilingual Apps, Using the Microsoft Multi

The majority of apps today are created for a single language audience, and mostly English. That is rapidly changing as we see the global expansion of app development on fire in markets most of us recognize as very significant. Our perspective is that this approach misses a substantial market potential even within the target markets, that being the multilingual nature of the world today, one that crosses all geopolitical boundaries. A good example of this is New York City, where English is still the most commonly spoken language, its cultural diversity is seen in a number of ways:

- 36% of population is foreign born
- 20% of the population speak little or no English
- Home to the largest Jewish community outside Israel
- Almost 25% of American Indians reside in the city
- 15% of all Korean Americans
- …

While New York City is the most culturally diverse city in the USA, it represents a fact across the country that every municipality of every size has a multilingual population.

Microsoft is committed to making access to technology across languages easier, working to lower the barriers where and whenever possible. In this context, we have created a set of tools and services that substantially lower the cost of entry into multilingual app development to enable developers of all types, Windows Desktop, Modern Store or Windows Phone, to quickly create multilingual apps for each of those platforms, but also enable easy workflow management that spans them by using the Multilingual App Toolkit and the integrated language services provided by MS Translator and the MS Language Portal.

This tutorial will take you through the process of setting up an app for proper globalization and localizability and then through the few steps needed to add languages, create machine aided translations, hand off strings to language specialists and merge them back into the project and to recycle across apps to reduce the costs just as we do on our products at Microsoft. We will demonstrate how you leverage the translated strings we have shipped in our products, created by professional translators as well as super quick machine translations from the Translator service, all at no cost to you.

This tutorial is for anyone who creates or wants to create apps and understands that they can reach more customers by adding languages. Whether you are just beginning or running an Enterprise development team, there is good stuff in this session for you.

Presenter:

Martin J. Dürst
Professor, Aoyama Gakuin University

Track 3 - Character Equivalences, Mapping, and Normalization

The wealth of characters in Unicode means that there are many ways in which characters or strings can be equivalent, similar, or otherwise related. In this tutorial, you will learn about all these relationships, in order to be able to use this knowledge for dealing with Unicode data or programs handling Unicode data.

Character relationships and similarities in Unicode range from linguistic and semantic similarities at the 'top' to equivalent representations in different character encodings and Unicode encoding forms at the bottom, with numerical and case equivalences, compatibility and canonical equivalences, and graphic similarities in the middle. The wealth of equivalences and relationships is due to the rich history of human writing as well as to the realities of character encoding policies and decisions.

Each of these relationships are ignorable in some processing contexts, but may be crucial in others. Processing contexts may range from use as identifiers (e.g. user ids and passwords) to searching and sorting. For most of the equivalences, data is available in the Unicode Standard and its associated data files, but the use of this data or the functions provided by various libraries requires understanding the background of the equivalences. When testing for equivalence of two strings, the general strategy is to normalize both strings to a form that eliminates accidental (in the given context) differences, and then compare the strings on a binary level.

The tutorial will not only look at officially defined equivalences, but will also discuss variants that may be necessary in practice to cover specialized needs. We will also discuss the relationships between various classes of equivalences, necessary to avoid pitfalls when combining them, and the stability of the equivalences over time and under various operations such as string concatenation.

The tutorial assumes that participants have a basic understanding about the scope and breadth of Unicode, possibly from attending tutorials earlier in the day.

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

09:15-10:00	絵文字：🏰 , 🎁 , and 🚀

Presenter:

Dr. Mark Davis
Unicode President and Co-Founder

Dr. Davis will discuss where emoji came from, why they have become so popular, where they've missed the mark, and what the future will bring.

In a few short years, emoji characters have become popular worldwide. This presentation will cover some of the history of emoji, describe the current emoji characters in Unicode, cover some of the deficiencies that people see in the current emoji (such as the lack of human diversity, and the missing HOT DOG), summarize some of the additions coming in new Unicode versions, and answer some of the most common questions about emoji.

10:00-10:30 - Morning Refreshments

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

Presenter:

Addison Phillips
Globalization Architect, Lab126 (Amazon)

Track 1 - 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 focuses on the activity of the W3C Internationalization Working Group, exploring developments over the past year, as well as the status and challenges of on-going work.

Presenter:

Nick Patch
Software Engineer, International, Shutterstock

Track 2 - Unicode Regular Expression Engines

Regular expression engines in most modern programming languages and libraries have been rapidly adding Unicode features in recent years. At Shutterstock, along with most other companies, we use a variety of programming languages, so it's important to know each language's strengths, weaknesses, and differences.

This talk will review Unicode regex features and compare support for these features in many popular engines as of 2014. Features discussed will include escape sequences, character properties, character classes, grapheme clusters, boundary anchors, letter casing, and normalization. Languages with core regex engines including Perl, Python, Ruby, Java, and JavaScript will be compared along with the ICU and PCRE libraries.

Presenter:

Ken Lunde
Senior Computer Scientist, Adobe Systems Incorporated

Track 3 - Developing & Deploying The World's First Open Source Pan-CJK Typeface Family

There is nothing extraordinarily new about Pan-CJK fonts, which are designed to serve multiple East Asian languages and regions using a single font resource, and whose development simply requires a lot of coordination and resources. What is extraordinary about Source Han Sans and the Google-branded Noto Sans CJK is that they represent the world's first open source Pan-CJK typeface families. Offered in a broad range of weights, these typeface families serve the needs of the East Asian community, and also comes with the right price tag: free. This presentation will detail what went into the development of Source Han Sans and Noto Sans CJK, including some of the pitfalls that were encountered along the way, along with the reasons behind the various deployment formats in which they are offered.

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

Presenter:

David Mohr
International QE, Adobe Systems Incorporated

Track 1 - Creating Software for Languages and Culture

Most software applications are designed and built in North America, catering to English-speaking, North American users. Ethnography is commonly used to ensure the product addresses the needs of domestic customers, but there is no analogous research with regard to foreign consumers. As a result, often times the features in these applications fail to sufficiently address the needs, expectations, and workflows of non-English-speaking users. These consumers are stuck with a crippled solution and little recourse. Although software engineering manuals minutely detail the technical aspects of bringing an English application to a foreign market and further touch on common design pitfalls, there are no recognized guidelines for how to most effectively design a product to address the requirements and needs of foreign customers from the outset. How do you create features tailored to the linguistically-defined and culturally-informed workflows of such users? This research follows a concerted attempt on the part of Adobe's Photoshop development team to create innovative designs specific to Middle Eastern customers, particularly Arabic- and Hebrew-speaking users. Software engineers develop complex applications, but their cultural lenses often prevent them from perceiving the expectations of unfamiliar customers. As part of the engineering effort, the author consciously incorporated a number of anthropological techniques not commonly applied to designing software for non-domestic consumers, particularly ethnography and validation by a distributed structured network of foreign users. There is significant potential for improvement in the design and development process and anthropology appears well suited to enhance software internationalization.

Presenter:

Martin J. Dürst
*Professor, Aoyama
Gakuin University*

Track 2 - Design Considerations for Internationalization in Ruby 2.2

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. This presentation discusses how to add internationalization functionality such as Unicode normalization, case conversion, and number formatting to Ruby in a true Ruby way.

Since Ruby 1.9, Ruby has a pervasive if somewhat unique framework for character encoding, allowing different applications to choose different internationalization models. In practice, Ruby is most often used with UTF-8.

Support for internationalization facilities beyond character encoding is available in various external libraries, but not yet in the Ruby core. As a result, libraries and applications may use conflicting and confusing ways to invoke internationalization functionality.

To use case conversion as an example, Ruby comes with built-in methods for upcasing and downcasing strings, but these only work on ASCII. An internationalization library may add separate functions for case conversion for the whole range of Unicode characters.

We study the interface of internationalization functions/methods in a wide range of programming languages and Ruby libraries. Based on this study, we propose to extend the current built-in Ruby methods with additional parameters to allow language-dependent, purpose-based, and explicitly specified functionality, in a true Ruby way.

This presentation is intended for users and potential users of the programming language Ruby, and people interested in internationalization of programming languages and libraries in general.

Presenter:

Alolita Sharma
*Director of Engineering,
Wikipedia*

Track 3 - Delivering Web Fonts at Wikipedia Scale

Wikipedia supports almost 300 languages for its multilingual content communities. As non-Latin language content grows exponentially, a breakthrough technology of delivering webfonts on demand has been deployed across 900 Wikimedia sites. This talk discusses user benefits derived, performance and scalability improvements made to deliver webfonts at Wikipedia scale.

Webfonts are fonts which are optimized for Web browsers. The challenges at scale include delivery of large font payloads which can exceed a megabyte and delays in transit to the client browser (e.g., slow phones or remote tablets). They also include deciding which fonts may be required for viewing potentially mixed content. This talk will discuss how Wikipedia is developing solutions for delivering webfonts at scale.

Wikipedia optimizations will be discussed including use of the Autonym font, font subsetting, tofu detection algorithms and context detection to help determine fonts needed.

Wikipedia provides an excellent platform for improving webfont delivery strategies that all Web-based content consumers, e.g., phones, tablets and desktops, need. The Wikipedia solutions are still bleeding-edge at scale but provide a solid starting point. Ultimately the challenge for Wikipedia and the Web is to seamlessly support over 6000 languages with many thousands of webfonts.

12:30-13:30 - LUNCH

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

Track 1 - Global Companies Adopting Quality Globalization

Presenters:

Adam Asnes
*President & CEO,
Lingoport*

Gary Lefman
*Internationalization
Architect, Cisco*

Michael Kuperstein
*Localization Engineer,
Intel*

Xiang Xu
*Senior Globalization
Developer, PayPal Inc.*

When adding best-practice quality tools to a large company software development process, ad hoc tools may need to be modified or processes modified to suit the tool. In this presentation, teams from PayPal, Intel, and Cisco will present the process they went through in evaluating, implementing, and improving static analysis tools to help detect and correct internationalization issues in enterprise software.

Track 2 - Complex Script Nuances and Challenges Inside Portable Document Format (PDF)

Presenter:

Mukul Sharma
*Computer Scientist,
Adobe Systems
Incorporated*

Text handling for RTL and Indic languages require complex script support due to the character properties or representation for these languages. Any text authoring tool which provide support for these languages works with complex script text engine. But what are complex script nuances and challenges for enablement of these languages? In this session I'll answer to this question by editing complex script PDF using Adobe Acrobat as an example and also discuss about challenges and solutions for enabling complex scripts

editing inside PDF format. Now we can edit the PDF multilingual content without depending on the source of the content. If you are new to complex scripts or planning for RTL or Indic languages enablement in your product then this session would be helpful for you.

Presenters:

Behdad Esfahbod
Internationalization Software Engineer, Google Inc.

Roozbeh Pournader
Internationalization Engineer, Google Inc.

Track 3 - Unicode, OpenType, and HarfBuzz: Closing the Circle

Unicode has always avoided providing too much font-related advice, and the OpenType specifications have always been ambiguous on how fonts should support certain complex scripts or harder cases. Font designers have been left with a hodgepodge of hard-to-decipher platform incompatibilities. With HarfBuzz, we planned to support every script, every character, and as many fonts as possible while complying with standards and remaining compatible with existing solutions. That's why we tried to close the loop by going back to Unicode and OpenType communities, in order to help international users have a platform-independent experience. We will report our recent achievements in collaboration with other platform developers, and share our goals for a more compatible future.

14:30-15:20	SESSION 4

Presenters:

Erwin Hom
Internationalization Engineer, PayPal Inc.

Michael McKenna
Globalization Engineering Leader, PayPal, Inc.

Track 1 - Using GPS to Track Your Position and Trajectory

So, suppose you've been asked to create a software application that can be easily released globally, where do you begin?

How do you track the process of the global-readiness of your application?

Whether you're creating an application for a few locales or a few dozen, you'll need some kind of system to measure its progress and success.

At PayPal, we've developed a Globalization roadmap for our development teams to follow and a system to track and grade the maturity of our products.

In this talk, we'll present the Globalization Maturity Model (GMM) which divides i18n capabilities into groups providing a roadmap of i18n support to be developed.

We'll also present the Globalization Product Scorecard (GPS) which can be used to track and grade the maturity of global-readiness in your products.

The GPS at PayPal is composed of three measures: the level of i18n maturity according to the GMM, a globalization quality score, and a global product management score.

Presenter:

Markus Scherer
Unicode Software Engineer, Google, Inc.

Track 2 - 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 53 & 54, including the latest support for Unicode 7.0 and CLDR 25/26, a new collation implementation, formatting of measurement units and durations, and other changes (see http://site.icu-project.org/download). The presentation will also touch on ICU's planned direction for future releases.

Presenter:

Pravin Dinkar Satpute
Senior Software Engineer, Red Hat

Track 3 - Project for Creating Efficient, Effective, Standard and Reusable Open Type Tables for Complex Script

Font development for complex script has been always remained a challenge due to involvement of different domains including Type design, Linguistics and Technical knowledge. Type design relates to the artistic domain, linguistics relates to peculiar knowledge regarding script and last but not the least technical knowledge relates to the Open Type specification and Apple Advanced Typography.
The effect of these challenges is easy to observe, we have millions of fonts for Latin script but when we consider the number of fonts for complex scripts including Indic we have a very limited number of fonts available.

In the last couple of years Unicode has added a number of complex script blocks and to get more and more fonts for these script it is very important to help the font development communities by removing unnecessary complexities in font development.

Lohit2 is a project targeting to remove the unwanted complexities from the font development process for complex scripts and to help type designers to quickly develop fonts. Lohit2 tables are effective (works on Uniscribe and Harfbuzz) and efficient (used minimum number of glyphs and compact Open Type tables).

15:20-15:50 - Afternoon Refreshments
15:50-16:40	SESSION 5

Presenter:

Murray Sargent
*Software Design
Engineer, Microsoft*

Shawn Steele
*Globalization for
Windows, Microsoft*

Track 1 - BiDi Internationalized Resource Identifiers

The Internationalized Resource Identifiers (IRI) is a generalization of the Universal Resource Identifier (URI) that permits many non-ASCII characters, such as most alphabetic characters and Chinese characters. Complications occur when right-to-left (RTL) characters are used in IRIs especially when displayed in RTL contexts. As the IRI reference discusses, use of the Unicode Bidi Algorithm is consistent in the way such IRIs are displayed in plain text, but some RTL IRIs are nearly unreadable. This talk discusses several ways to render them legibly and recommends that the directionality of the URL neutral characters be given by the paragraph directionality.

Presenters:

Tex Texin
*Globalization Architect,
Xencraft*

Craig R. Cummings
*Principal Software
Engineer -
Internationalization,
Informatica*

Track 2 - Comparing JavaScript Libraries

Which JavaScript library is best for international deployment? This follow up to last conference's session presents the results of further 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: Dojo, JQuery, and YUI as well as Closure, Microsoft, and the ECMA-402 work. The results still surprise and will be useful to anyone designing new international or multilingual JavaScript applications or supporting existing ones.

Presenters:

Andrew Glass
*Program Manager,
Microsoft*

John Hudson
*Co-Founder, Tiro
Typeworks Ltd.*

Track 3 - Shaping in the Post-tofu Era

The rapid progress of font developers towards the goal of providing fonts that cover all of the code points in Unicode means that tofu ( ) will soon be banished. This is a significant milestone for the Unicode community but it is not the end of the road. Alongside character support, there must also be shaping support. It is not enough to display a glyph for every character, but they must also be displayed, joined, positioned, ligated, etc., correctly. To accomplish that, shaping engines provided by platforms must provide basic shaping support required for all writing systems that font developers can leverage. In turn, font developers must know how to interact with these engines to develop fonts that meet the needs and expectations of the user community.

This talk will provide a brief history of shaping engines on Windows and describe the principles that have been guiding evolution of shaping on the Windows platform, a look at the impact of recent results from efforts to standardize shaping behavior (see Unicode, OpenType, and Harfbuzz) on font developers and describe with a demo, how to develop OpenType layout for a previously unsupported script.

16:50-17:40	SESSION 6

Track 1 - The Upcoming Changes and Impacts from Email Address Internationalization

Presenters:

Joseph Yee
*Technical Account
Manager, Afilias*

How email address being used changse a lot. The email address evolves from contact identifier to login/credential identifier. Many services and applications treats email address as user's primary key. The format of email address may not change much until recently, the introduction of new gTLD in IDN and the new email standards. The new email standards allow characters beyond non-accented Latin characters in both username part (local part) and domain part. All parts use Unicode and UTF-8 encoding, however, each part adopts different rules. This presentation discusses the changes coming from the new standard, and what impacts it may bring to applications development and solution providers who utilize email address as identifier.

Presenter:

Murray Sargent
*Software Design
Engineer, Microsoft*

Track 2 - UTF-8 RTF

Word's RTF format is the oldest widely used rich text format. It is the only format that can travel through time: older apps can read RTF from newer apps and vice versa. It has a simple syntax analogous to XML. But it is a bit mired in the old code page world in that most common Unicode characters are converted to code-page values. This talk shows several ways how RTF can be written using Unicode with no reference to code pages. In particular, this allows one to have a light-weight rich-text format that includes language information.

Presenters:

Qin Lu
*Professor, Hong Kong
Polytechnic University*

Track 3 - Supporting Chinese Character Variants in Hong Kong through Ideograph Variation Sequences

This talk will introduce an ongoing project in Hong Kong that makes use of the Ideographic Variation Sequence (IVS) and the associated Ideographic Variation Database (IVD) developed by the Unicode Consortium for character glyph registration. Hong Kong uses the traditional Chinese writing system similar to that of Taiwan and thus used the Big5 encoding for many years. But, Chinese characters used in Hong Kong do have different variant glyphs. The current CJK computer coding for Chinese characters is done at the character level. However, a single Chinese character may still have different glyph shapes called variants. Variants normally do not change the meaning of the character, yet, if coded separately, will cause problem in searching and indexing. Thus, in the processing of ISO 10646/Unicode standardization, a

so called unification process for the ISO10646/Unicode standard can cause confusion and inconvenience in certain applications. There are applications where different written styles of the same logical character may need to be included in the same document requiring these variants to be separately encoded and specified at the character level. The Ideographic Variation Sequence (IVS) and the associated Ideographic Variation Database (IVD) developed by the Unicode Consortium for character glyph registration are quite suitable for Hong Kong's Chinese variant specification. This talk will explain how the IVS technology is used to encode these Hong Kong specific Chinese variants, the process of the review, the production of variant, and the production of the Hong Kong Character variant Specification, and the registration in the IVD of Unicode.

18:00-19:00 - CONFERENCE RECEPTION

Wednesday, November 5, 2014

09:00-09:50	SESSION 7

Presenter:

Addison Phillips
*Globalization Architect,
Lab126 (Amazon)*

Track 1 - Time Out of Joint

Time values are fundamental to many software applications--from keeping your calendar up to date to identifying awards in a game, from logging events locally to scheduling meetings around the world. Working with events, time zones, and time zone independent values ("floating dates") can be confusing.

This session, from the editor of W3C I18N Working Group's note "Working with Time Zones", reviews how time works in many computing systems, looks at different use cases for working with time, and provides some guidance on how to structure data to get just the right time at just the right time.

Presenters:

Erwin Hom
*Internationalization
Engineer, PayPal Inc.*

Aarti Ashok
*Product Architect,
Globalization
Technology, PayPal, Inc.*

Track 2 - Address Standards, Formats, and Nightmares

If your application deals with addresses for multiple countries, this talk will highlight the challenges in supporting address in an internationalized application. In this talk, we'll present:

- Variations of Address Formats
  - Complex Address Format in UK, Brazil, Japan, and others.
  - Street Complement component in the Indian Address Format
  - Reading an Address as a Native versus reading it as a Non-Native (for example, English speaker)
  - Providing for local and international formats to help in cross-border trade
- Handing address formats in the UI - address entry forms
- Handling address data in the app (This is an opportunity to present the Canonical Format, Locale Neutral Format)
- Devising an XML layout syntax based off CLDR and LDML to put name and address layout meta data in a machine-readable form

Presenter:

Marc Durdin
CEO, Tavultesoft Pty Ltd

Track 3 - The Future of Keyboard Input on Touch Devices

Much work has already been done with keyboard-style input on touch devices, using techniques such as touch-and-hold popups, automatic corrections and suggestions, and companies such as Swype working to improve both accuracy and efficiency of input through gesture-based input.

However, once we move away from Latin-based text and Far Eastern script input methods, many languages and scripts have only rudimentary layouts available, mostly translated from existing, hardware-limited layouts. We have only scratched the surface of how touch layouts could be improved, particularly when coupled with intelligent input software such as Keyman. I will walk through some experimental keyboard layouts that we have developed, for a number of languages, including Lao, Tamil, Amharic, Ancient Greek, Egyptian Hieroglyphics and Arabic, where we have attempted to break free of the design limitations and thinking imposed by desktop QWERTY keyboard layouts.

10:00-10:50	SESSION 8

Presenter:

Ken Lunde
*Senior Computer
Scientist, Adobe Systems
Incorporated*

Track 1 - Building Source Han Sans & Noto Sans CJK

For the benefit of those who want to know more about the nuts and bolts of building Source Han Sans and Noto Sans CJK, the world's first open source Pan-CJK typeface families, this highly-technical presentation dives into the actual build processes and techniques that were used, starting with the raw glyph data from the four participating type foundries, and ending with installable font resources.

Presenters:

Rafael Xavier de Souza
Team Member, jQuery

Track 2 - jQuery Globalize ♥ CLDR

jQuery has changed the way that millions of people write JavaScript, believing in a world in which all web content is built on open standards and is accessible to all users.

Learn how jQuery is improving its internationalization library - Globalize - to leverage the official CLDR

JSON data, allow users to load as much or as little data as they need, avoid duplicating data if using multiple i18n libraries that leverage CLDR, and that run in browsers or node.js.

Presenter:

Craig R. Cummings
Principal Software Engineer - Internationalization, Informatica

Track 3 - Bidi on Android and iOS

This session is will be an overview of the how-tos for developing bidi responsive web design (RWD) applications as well as developing native bidi Android and iOS applications. In addition to development techniques, some hints, tips, and tricks will be covered.

10:50-11:10 - Morning Refreshments

11:10-12:00	SESSION 9

Track 1 - Expanding the Unicode Repertoire: Un-encoded Scripts of Africa and Asia

Presenters:

Deborah Anderson
Researcher, Dept of Linguistics, UC Berkeley

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

This talk will highlight trends and issues related to scripts from these regions through several case studies, and offer possible strategies on how to make progress on the remaining scripts. African and East Asian scripts will be discussed by Dr. Deborah Anderson, director of the Script Encoding Initiative, while South, Central, and South-East Asian scripts will be presented by Anshuman Pandey of the University of Michigan.

Track 2 - New in CLDR Locale Data

Presenter:

Mark Davis
Chief Internationalization Architect, Google Inc.

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 overview of CLDR, then focuses on recent and forthcoming enhancements, including extended plural support, coverage improvements, keyboard data, the data collection process, and changes introduced in CLDR 26 (planned for 2014-9-15). A significant amount of time will be reserved for demos of behavior based on CLDR data.

See also http://cldr.unicode.org/index/downloads/cldr-25

Track 3 - Mobile I18n Testing Toolbox

Presenter:

Katsuhiko Momoi
Staff Test Engineer, Google, Inc.

For the past year we been developing a set of APIs/tools that we recommend as the standard testing tools for Mobile apps. This session introduces the audience to the Mobile I18n Testing Tolbox at Google. Many Google apps are localized into nearly 60 languages and yet resources to test the international quality of these apps are never sufficient. We have developed a set of tools/APIs/recommendations for mobile i18n testing that help our teams in providing i18n test coverage with limited testing resources. In this talk, we discuss these tools (Mobile I18n Testing Toolbox) with focus on how to use them on the Android platform. The tools include Multi-locale Switches, Intl Sanity Checker -- an easy to use comprehensive locale format test API, Keyboard/IME App/Test API that checks keyboard compatibility against a given input box in an app, new pseudolocalizer integrated into Android Open Source Program, and UI breakage test tool with a crawler for Android. We are in the process of open sourcing most of these tools/APIs and hope to report on the progress of that effort as well.

12:00-13:00 - LUNCH

13:00-13:50	SESSION 10

Track 1 - Unicode Initiatives in the Middle East

Presenter:

Thomas Milo
MARA Representative- DecoType Partner, MARA - DecoType

The Sultanate of Oman is setting up a new institute for Unicode Research and Development in the widest sense of the term, with a focus on the whole region, not limiting itself to Arabic and RTL issues.

Track 2 - CLDR Users' Panel and Discussion

Moderator:

Steven R. Loomis
Software Engineer, IBM

This panel discussion focuses on the experience of direct users of CLDR data (primarily users other than ICU) including how they use the data and what issues have been encountered using LDML and CLDR data. This discussion will also include ample time for general Q&A about CLDR.

Track 3 - Top 10 Best Practices on Internationalizing a Start-up

Presenter:

Claudia Galván
Technical Advisor, Early Stage Innovation

Startups have great challenges and opportunities, they are small, dynamic and flexible. They also face limited resources, time constraints and fierce competition. When startups start looking at additional revenue and growth opportunities, some look into expanding into international markets. There is an art of pivoting successfully into a global player. I have been working with startups and I have gather best practices to help them succeed in this process.

13:50-14:40	SESSION 11

Presenter:

Saurav Gupta
Sr. Globalization Lead, Adobe Systems India Pvt. Ltd.

Track 1 - Market Trend of Indic Languages and Their Challenges

India is always an emerging market for all domain and functions, but often we tend to miss out on things next to you. That's what has happened with Indic languages. With large number of user's base, still these languages were ignored for long but not anymore. The market has seen a steep rise and these languages are making their way in the Localization World. Adobe has started exploring this untouched area, started with products like InDesign, Illustrator, Adobe TV and so on.

These languages have their own peculiar characteristics that stand them different from each other and cause challenges while making our products ready for the market, starting from character issues to word formations.

This will cover the entire spectrum of Market Trends, Languages and dialects, and mitigation for extending the support for INDIClanguages in Adobe Products. Let's explore these languages, their marketing (financial and user prospects) trends, interesting revelations, testing challenges and answers to some that holding you up from moving on.

Presenter:

Steven R. Loomis
Software Engineer, IBM

Track 2 - CLDR Users' Panel and Discussion (Cont.)

This panel discussion focuses on the experience of direct users of CLDR data (primarily users other than ICU) including how they use the data and what issues have been encountered using LDML and CLDR data. This discussion will also include ample time for general Q&A about CLDR.

Presenter:

Alolita Sharma
Director of Engineering, Wikipedia

Track 3 - Internationalization Testing at Wikipedia

14:50 – 15:10 - Afternoon Refreshments

15:10 - 16:00	SESSION 12

Presenter:

Manish Aggarwal
Computer Scientist, Adobe Systems Incorporated

Track 1 - Enabling Indic Support in Applications

Since CS6, Adobe is providing support for Indic text layout in its text editing products. Indic scripts (such as Devanagari) are complex scripts and are therefore non-trivial to implement. For example, the input sequence order of characters can be different from the actual pictorial representation, and they rely heavily on diacritics and ligature formation. This talk will cover requirements, challenges, and architecture of the Indic text shaping engine for many Indian languages.

India is huge market with high revenue potential for application development companies. Recently Adobe started supporting Indic languages in its products such as InDesign, Photoshop and Illustrator. But supporting complex Indic scripts such as Devanagri, Malayalam, Tamil and many others is non-trivial task and is full of challenges. Ex, in Indic scripts, input sequence order of characters can be different from the actual pictorial representation. Thus, reordering of characters is required before processing the text. Also, there could be thousands of syllable (user perceived single unit of a character) in an Indic language which heavily depends on the diacritics and ligature information from the font for their correct shaping. Thus, a dedicated text shaping engine is required to enable Indic languages in applications, which should be responsible for the correct shaping of text. I plan to discuss basics of Indic languages and the architecture behind their text shaping, which will be mainly focused around the following:

- Challenges posed by Indic scripts and how Unicode helped in handling them.
- High level architecture details of Indic text shaping engine
- Algorithm and rules for parsing Indic scripts

Presenter:

Yoshito Umaoka
Senior Software Engineer, IBM

Track 2 - Chronologies in the Non-Western World

Although there are many calendar systems available in the world, the Western (Gregorian) calendar is almost exclusively used in software industry. These local calendar systems are used for settings dates for traditional or religious events in non-Western countries. In some countries, official documents must use their local calendar dates.

Some existing programming platforms and common libraries, such as Microsoft Windows and ICU, have been supporting non-Western calendar calculation. However, majority of software running on these

runtime environments only support Western calendar, especially when calendaring and scheduling is involved.

This session will discuss the requirements for Non-Western calendar in the world, the challenges for supporting non-Western calendar systems in software applications, and the technologies currently available for Non-Western calendar calculation and display, including new date and time APIs (JSR-310) in Java 8. The session also touch on a new draft proposal (iCalendar/RSCALE) for calendaring and scheduling standard specifying events in non-Western calendars.

Presenter:

Tex Texin
Globalization Architect, Xencraft

Track 3 - Agile Internationalization User Stories

User stories are the way that Agile Methodology describes the functionality of the software being developed.

Each story describes an action or need of a user and in so doing defines the functions the software must provide and the requirements it must satisfy.

This session will describe the mapping of an internationalization checklist into a suite of user stories that are used in internationalizing a software project.

16:10 - 17:00	SESSION 13

Track 1 - Advanced Font Issues Panel

Presenters:

Behdad Esfahbod
Internationalization Software Engineer, Google Inc.

Andrew Glass
Program Manager, Microsoft

John Hudson
Co-Founder, Tiro Typeworks Ltd.

Roozbeh Pournader
Internationalization Engineer, Google Inc.

This session will feature a round-table discussion with some short presentations and demos on the topic of complex font issues, from input to rendering.

Track 2 - Automated System for New Unicode and CLDR Globalization Features

Presenters:

Su Liu
Globalization Architect, IBM

Aya Elgebeely
Software Developer, IBM

Ashraf Gomaa
Software Developer, IBM

David Clissold
User Space Architect, IBM

Fan F Yang
Software Developer, IBM

Several automated tools for CLDR and Unicode implementation are discussed in this presentation. Enabling and maintaining globalization packages are an expensive, error-prone, and time consuming process in an operating system software development cycle. Automating tedious processes is the best practice in CLDR and Unicode work, and it can bring consistency and transparency to globalization features both on application and OS levels. In AIX, the globalization team is focusing on automated CLDR and Unicode support and inventing a set of useful toolkits for leveraging AIX globalization features. A CLDR locale builder is designed to robotically build and package a new locale based on the latest CLDR data. A Unicode development tool is created to automatically update related LC_CATEGORY files, and update all impacted locales to the latest Unicode version. In addition, a functional verification engine is formed to cover the most globalization features in AIX. This presentation also covers CLDR and Unicode problems and issues (such as POSIX locale compatibility, server-side locale integration, dual locale-object handling, legacy locale migration, and locale alias name policy) in operating system software.

Track 3 - Pseudo Localization in Action

Presenters:

Xiang Xu
Senior Globalization Developer, PayPal Inc.

Erwin Hom
Internationalization Engineer, PayPal Inc.

Michael McKenna

Pseudo localization is an effective way to detect i18n/l10n issues, the effective use of pseudo localization can mean huge savings for a company. A well designed flexible pseudo localization tool is critical. At PayPal, we use pseudo localization to help us process multiple content types (XML-based content, Java properties, mobile resource files, etc.) with various pseudo localization patterns on multiple platforms. In addition to supporting classic pseudo localization practices such as dynamic text expansion and readability of resulting text, it also supports pseudo patterns that are custom made for Asian languages as well as RTL languages. With this tool, it is easy to add support for future content types and more pseudo localization patterns, as well as integrating Machine Translation.

Globalization Engineering Leader, PayPal, Inc.

In this talk, we will present challenges and success of our pseudo-localization tool, the kind of issues that are detected by the tool, as well as the use of it as an education tool to engineering and UI/content designers. We will discuss the documented increase in quality since integrating the tool with development, design, and CI processes.

Program is subject to change.

- To Register for IUC38: 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.

[iuc38/_borders/bottom.htm]