

- home
- advisory committee
- program committee
- conference program
- hotel
- registration
- press room
- become a sponsor
- IUC 31 exhibits
- become an exhibitor
- past conferences
- send me more info

**Hotel cut-off:**  
**09/21/2007**  
**Early-Bird**  
**Registration by:**  
**09/05/2007**  
**Venue:**

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

## Program

**Monday, October 15, 2007**

<b>09:00-10:30</b>	<b>MORNING TUTORIALS</b>
--------------------	--------------------------

*Presenter:*

**Richard Ishida**  
*Internationalization  
 Activity Lead,  
 W3C*

### Track 1: An Introduction to Writing Systems & Unicode

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

*Presenter:*

**Addison Phillips**  
*Internationalization  
 Architect,  
 Yahoo!*

### 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 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. Some of the topics covered will include: character encodings and Unicode; processing text in different languages; preparing for the localization (translation) of user interfaces; making applications "locale-aware", including format and display differences; as well as approaches to delivering multi-lingual and multi-locale software or content.

*Presenter:*

**Vladimir Weinstein**  
*Software Engineer,  
 Google*

### Track 3: ICU4C in Action

International Components for Unicode (ICU) is a very popular internationalization software solution. However, similar to any complex product, a learning curve is involved. The goal of this tutorial is to help new users of ICU4C install and use the library. Topics include: Installation, verification of installation, introduction and detailed usage analysis of ICU4C'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 C and C++ programming. After the tutorial, participants should be able to install and use ICU4C for solving their internationalization problems.

10:30-10:45 - Morning Refreshments

<b>10:45 – 12:30</b>	<b>MORNING TUTORIALS (Cont'd.)</b>
----------------------	------------------------------------

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

*Presenter:*

**Doug Felt**  
*Google*

### Track 2 - Internationalization: An Introduction (Cont'd)

### Track 3 - Applied ICU4J

ICU4J is an open-source library for internationalization in Java. It is designed to be a 'drop-in' replacement/enhancement for Java APIs, providing more features, more data, and equivalent or better performance. This tutorial will show how to apply a number of ICU4J features with particular attention to differences between ICU4J and standard Java functionality.

12:30-13:30 - LUNCH

13:30-15:30

AFTERNOON TUTORIALS

*Presenter:*

**Asmus Freytag**

*President,  
ASMUS, Inc.*

### **Track 1 - Unicode 5.0 Tutorial: Fundamental Specifications**

The Unicode 5.0 Tutorial systematically presents the details of fundamental specifications that are part of the Unicode Standard. Topics include: organization of the Unicode code space; principles used to allocate and unify characters; encoding forms including definition of UTF-8, UTF-16, UTF-32 and when to use each; how to use byte order mark; combining characters and equivalent code sequences equivalent; format characters and other special characters and code points; organization of the Unicode Standard. This part of the Unicode tutorial is recommended for anyone interested in a systematic overview of the key aspects of the standard. Detailed technical or programming experience is not required.

*Presenter:*

**Pierre Cadieux**

*President,  
i18N Inc.*

### **Track 2 - Globalization: SQL Server vs. Oracle**

A head-to-head comparison of Oracle and SQL Server support for character sets, locales, collation, global stored procedures, etc. Learn that writing cross-platform database code with stored procedures is almost impossible. Thanks to the SQL standard, database queries and schemas are roughly compatible (e.g. VARCHAR vs. VARCHAR2, etc.). Even stored procedures have a decent measure of compatibility... in English, that is. But when you start considering global databases, with stored procedures that process global data on Oracle and SQL Server, compatibility is almost non-existent. This tutorial advances topic by topic: encodings, locales, collation, stored procedures. For each topic, the tutorial presents: Oracle features, SQL Server features and a comparative summary. You may discover that the road is rocky ahead!

*Presenters:*

**Elsebeth Flarup**

*Globalization Architect,  
Soeren Bendtsen  
Advisory IT Specialist,  
IBM Corp.*

### **Track 3 - Best Practices in Software Localization**

Software localization and internationalization are conceptually separate tasks, but they are best executed with full integration and interlock between the two. This tutorial uses practical examples to demonstrate how localization that is built into the process, starting from the design phase, may help lower your cost and improve time-to-market for localized versions. Topics include: Building international support into the product from the beginning; globalization verification testing and how to use pseudo localization effectively; dos and don'ts when creating translatable text; translation file formats; translation file check tools and how they can reduce translation problems, build issues and test duration; translation verification testing; source control and change freezes; terminology management and 'controlled English'; computer aided translation tools such as translation memory based systems; how to build localization project schedules and the interlocks required with development; tips on how to apply the concepts and techniques on projects where the translation process has been started in a non-optimal fashion. Demos will be used to illustrate tools and processes.

*Presenter:*

**Sayuri Wijaya**

*Program Manager,  
Erik Fortune  
Development Manager  
for MUI,  
Microsoft*

### **Track 4 - Writing Win32 Multilingual Applications Using the Windows Vista MUI Technology**

In Windows 2000, Microsoft introduced the Multilingual User Interface (MUI) technology, which enables users to change the display languages for the operating system from a list of available languages. Starting with Windows Vista, this MUI technology and a set of associated APIs are made available to Win32 application developers. This session is intended to introduce the benefits and capabilities of MUI in Windows Vista and to provide the necessary knowledge and best practices to use the MUI technology and its associated APIs to develop multilingual applications for Windows. The following topics will be covered: 1. The benefits of using MUI technology 2. Introduction to MUI technology in Windows Vista 3. How to use MUI technology in your application development 4. How to control the resource content in the Language Neutral and MUI files (resource configuration file) 5. How to use the rc.exe and muirct.exe tools to generate Language Neutral and MUI files 6. MUI API introduced in Windows Vista to take advantage of the UI language settings, and the best practices for using these APIs including down-level OS support 7. Step-by-step 'gray-form' demo of developing multilingual app using MUI technology

15:30-15:45 - Afternoon Refreshments

15:45-18:00

AFTERNOON TUTORIALS

*Presenter:*

**Asmus Freytag**

*President,  
ASMUS, Inc.*

### **Track 1 - Unicode 5.0 Tutorial: Unicode Algorithms**

The Unicode Standard and related specifications by the Unicode Consortium specify a number of algorithms. The specification of these algorithms in the Unicode Standard depends on the Unicode Character Properties. This part of the Unicode 5.0 Tutorial surveys the algorithms specified in the Unicode Standard, and extends the discussion of Unicode character properties as they relate to each algorithm. It covers many general aspects of Unicode algorithms: Unicode Algorithm and the difference between an abstract algorithm from an actual implementation; relation between algorithms and Unicode Character Properties; techniques to access character properties. Several algorithms are discussed in more detail for example: Unicode Normalization and the requirements it addresses, including a discussion of the Unicode Normalization forms NFC, NFD, NFKC, NFKD, their interaction with the Web and what programmers need to know in applying normalization; the Unicode Bidirectional Algorithm, and its interaction with text layout; text boundary determination and character foldings and much more. This part of the Unicode 5.0 Tutorial is more detailed and will touch on the description of

algorithms and other material that may require some familiarity with technical concepts.

*Presenter:*

**Pierre Cadieux**  
*President,  
I18N Inc.*

**Track 2 - Making Sense of Oracle Character Sets and Length Semantics**

Everything you need to know to work with Oracle character sets. A new model of Oracle character sets is presented, involving five character sets: database, national, client, and more! The model is mapped to Oracle usage in C/C++/Java/.NET. It is then used to explain the subtleties and pitfalls of Oracle transcoding. Numerous transcoding scenarios are illustrated visually with the model, as are the various parameters controlling SQL literal transcoding and Oracle's "form-of-use". Length semantics are then introduced along with the related SQL and PL/SQL functions. Finally, with all these features understood, the presentation finishes by discussing the pros and cons of the various ways of implementing Unicode in Oracle.

*Presenter:*

**Tex Texin**  
*Internationalization  
Architect,  
Yahoo!*

**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, XHTML, XML (eXtensible Markup Language; for general markup), and CSS (Cascading Style Sheets; for styling information) are given particular emphasis. The tutorial addresses language identification and selection, character encoding models and negotiation, text presentation features, and more. The design and implementation of multilingual Web sites and localization considerations are also introduced.

*Presenter:*

**Douglas R. Davidson**  
*Software Engineer,  
Apple, Inc.*

**Track 4 - Extending Mac OS X's International Support**

Mac OS X ships with extensive international support, but it also has a rich set of plug-in architectures that allow third parties to supply additional features. This tutorial provides a hands-on discussion of the localization and internationalization architecture of Mac OS X from a developer perspective, and shows how to create new input methods, keyboard layouts, locales, fonts, text services, and other components useful in extending international support on Mac OS X. Detailed examples will be presented. The discussion will be relevant to all versions of Mac OS X, but particular attention will be paid to Mac OS X Leopard.

**Tuesday, October 16, 2007**

09:00-09:15	<b>WELCOME &amp; OPENING REMARKS</b> <b>Mark Davis</b> - President, Unicode Consortium
09:15-10:00	<b>KEYNOTE – Graphic Speech and Graphic Song</b> <b>Robert Bringhurst</b> , Poet, Typographer, Linguist and Cultural Historian  Humans have been translating human language into graphic form for at least 5,000 years and have built up a lot of sophisticated resources in that time. Many of these resources are now meticulously itemized in Unicode. This is a great help. But whatever we do in language, we are always just beginning. When we speak, we make a sequence of speech-sounds, but we also sew them together into a shape. If the shape is lyrical enough, we say we've crossed the boundary that separates speech and song. It's much the same with written and printed messages. Making or finding the right characters is a start. Then we assemble them into a shape. And the way we make or choose them gives them style. Silent though they are, sometimes they fit together so well they seem to speak. Sometimes, in fact, they seem to sing. But even that is just a beginning.
<b>10:00-20:00 - EXHIBIT AREA OPEN</b>	
10:00-10:30 - Morning Refreshments in Exhibit Area	
10:30-11:20	<b>SESSION 1</b>

*Presenters:*

**Naoto Sato**  
*Java I18n Engineer,  
Sun Microsystems*  
**Craig R. Cummings**  
*Principal Software  
Engineer,  
Oracle*

**Track 1 - New Internationalization Features of the Java Platform**

See what internationalization features are in the present version and planned for the next version of the Java Platform -- Java SE 6 and 7. The talk will cover the existing features in Java SE 6, such as Locale Sensitive Services SPI, Normalizer API, ResourceBundle enhancements, and new Japanese calendar support. Then it will cover what will be in the upcoming JDK 7 release.

*Presenter:*

**Mark Davis**  
*Google,  
(President, UNICODE  
Consortium)*

**Track 2 - Unicode in Google**

Google makes extensive use of Unicode in all of its products. For example, all web pages -- no matter what their original encodings -- are mapped to Unicode for processing. This presentation will discuss some of the uses of Unicode in various Google products, and some of the challenges involved in processing Unicode on an extremely large scale. It will also discuss some of the approaches to internationalization that have been found to be particularly effective.

---

*Presenter:*

**Vladimir Weinstein**  
*Software Engineer,  
Google*

### **Track 3 - ICU on a Diet**

International Components for Unicode is a full-featured internationalization library. It includes frameworks for dealing with all important internationalization tasks, as well as a very extensive set of supporting data. The completeness comes with a price – ICU libraries tend to need a non trivial amount of space. This talk discusses several ways to reduce the size and customize an ICU installation, such as choosing the appropriate feature set for the application and reducing the size in installed data. Both Java and C/C++ libraries will be discussed.

---

*Presenter:*

**Jiangping Wang**  
*Assistant Professor,  
Webster University*

### **Track 4 - Internationalization in Computer Science Curriculum**

Traditionally computer science curricula in US universities are designed to teach students theories, design, and implementation in English only. However, without internationalization consciousness many design and programming tasks can disrupt the application's ability to function globally. Students need to be exposed to internationalization concepts and practices so that they can expand the knowledge after entering the industry. This presentation will discuss opportunities and challenges in integrating internationalization in computer science curriculum and teaching internationalization in computer science courses. The discussion will investigate challenges that are preventing us from integrating internationalization contents into curriculum and the approaches to tackle them.

---

**11:30-12:20**

**SESSION 2**

*Presenters:*

**Jim DeLaHunt**  
*Principal,  
Jim DeLaHunt &  
Associates*  
**Daniel Strebe**  
*Adobe Systems*

### **Track 1 - SING "gaiji" Architecture in Adobe Creative Suite 3**

Writers in Chinese, Japanese, and Korean (CJK) languages draw from an infinite collection of Chinese-derived characters, and only some are in fonts. Those that aren't, are known as "gaiji." Learn why gaiji are important. See the SING Gaiji Architecture, in Adobe's Creative Suite 3. Learn how SING extends your CJK fonts with individual new OpenType-based "glyphlets". Embedded in documents, glyphlets move through the workflow. Consider SING's implications for the Unicode's character-glyph model and Ideographic Variation Sequences and for anyone with text in CJK languages, be it for publishing, for corporate databases, or for the web and cell phones.

---

*Presenter:*

**Michael Kaplan**  
*Technical Lead,  
Microsoft*

### **Track 2 - Embedding & Linking & Fallback, Oh My! (Getting the Characters You Want)**

Whether using Win32, the .NET Framework, or Windows Presentation Foundation, the battle to make sure that text will always display properly seems like a neverending one. This talk will review the different technologies used and will discuss the benefits and drawbacks of each. Many of the technical and legal issues that surround the problem of the proper display of Unicode text will also be covered.

---

*Presenter:*

**Stanislav Malyshev**  
*Software Architect,  
Zend Technologies*

### **Track 3 - Climbing the Tower of Babel with PHP**

The Unicode and i18n support in PHP continues to evolve. This talk will provide an overview of the most salient features of PHP 6's Unicode support and illustrate the new internationalization features with a variety of demos on topics such as: Character set conversion; Text boundary analysis; Working with international dates and calendars; Transliteration and text normalization; Working with character sets and properties.

---

*Presenter:*

**Elizabeth Pyatt**  
*Instructional Designer/  
Instructor in Linguistics,  
Penn State*

### **Track 4 - Moving a Large Scale University to Unicode Usage**

This presentation discusses efforts to facilitate Unicode usage at Penn State. University support is particularly challenging because the audience includes native speakers, language learners and monolingual English tech support staff. Efforts have included documentation, (<http://tlt.its.psu.edu/suggestions/international/>), researching Unicode in new technologies (e.g. blogs/Flash) and outreach to multiple departments to determine campus software needs (e.g. fonts/keyboards/text editors). The main lesson learned has been that "each language has its own story" and that users respond best when given specific details for their situation. Thus, outreach and documentation has been structured around specific languages and software, even if actual Unicode implementation is more general.

---

**12:30-13:30 - LUNCH**

---

**13:30-14:20**

**SESSION 3**

*Presenter:*

**Ned Holbrook**  
*Software Engineer,  
Apple, Inc.*

### **Track 1 - International Features of Mac OS X Leopard for Developers**

Mac OS X has long been an excellent platform on which to build Unicode-enabled applications and services. The latest release, Leopard, builds on the foundation of prior releases by offering a range of new and improved features for Unicode developers. This talk will provide an overview of these new features. The topics covered will include application programming interfaces (APIs) for collation, tokenization, dictionaries, input methods, locale data, fonts, and line layout.

---

*Presenter:*

**Pierre Cadieux**  
*President,  
i18N Inc.*

### **Track 2 - What your Boss Needs to Know about Internationalization**

If programmers are notoriously optimistic, sometimes the powers that be are on an entirely different planet: "After all, internationalization is just about translating a few strings! Let's do this on the same budget, same deadline and with no training. Oh, and with the same English-

only testing, of course!" This presentation is a visual and entertaining overview of the various issues that have to be dealt with when internationalizing software: layout issues, message formatting, character sets, input methods, text rendering, text processing, currency, calendars, searching, forms, colors, addresses, etc.

---

*Presenter:*

**Martin J. Dürst**  
*Associate Professor,  
Aoyama Gakuin  
University*

### **Track 3 - Internationalization of the Ruby Scripting Language**

Ruby is a purely object-oriented scripting language that is rapidly growing in popularity due to its high productivity. Because it was invented in Japan, some basic internationalization features are available, but there is still a lot of work to do. This presentation will give a short overview of the most important features of Ruby, and introduce the available internationalization-related features, concentrating on how to use Ruby with Unicode, which in Ruby's case means UTF-8. An outlook of planned directions for further internationalization work is also given.

---

*Presenter:*

**Richard Cook**  
*Linguist,  
UC Berkeley*

### **Track 4 - The Character Description Language (CDL) Digital Humanities Start-up**

The Character Description Language (CDL) Digital Humanities Start-up is a project to provide CDL software for the mapping of Chinese, Japanese, Korean, and Vietnamese (CJKV) script elements, for the augmentation of a standard database of CDL descriptions open to members of international standards bodies and to the public.. This presentation will describe the fundamentals of CDL, and outline the short- and long-term CDL project goals. In the short term, the CDL team seeks to tame CJK encoding; in the long term, we propose to build a collaborative tool for management and publication of all UCD glyph data, for CJK and beyond.

---

**14:30-15:20**

**SESSION 4**

---

*Presenter:*

**Russ Rolfe**  
*Sr. Program Manager,  
Microsoft*

### **Track 1 - Windows Vista Language Support — How Does it All Fit Together**

Microsoft's Windows Vista has 36 localized builds and 50 plus language interface packs (LIP) as well as supports 100's of different languages. The localized builds can come in many flavors -- Starter Edition, Home Basic, Home Premium, Business, Enterprise, and Ultimate. Besides the localized versions of Windows Vista, there is also the support for creating and displaying content in many different languages. This presentation will sort out the different types of and levels of language support that can be found in each of these versions and how they all relate to each other.

---

*Presenter:*

**Addison Phillips**  
*Internationalization  
Architect,  
Yahoo!*

### **Track 2 - Making Sense of Global Communities**

User created or generated content and the social networking phenomenon represents a new range of richness in user interaction with the Web. It also presents new challenges as communities designed for the domestic English marketplace extend globally and are used in other regions, cultures, and languages. Users like the idea of global reach, yet wish their content to remain relevant and accessible. Sub-groups may still organize themselves around cultural or linguistic similarities, but this need not be the defining factor in user organization. As social networks move toward the mainstream, designing them to work with languages, cultures, and international laws becomes more complex. This presentation discusses the challenges involved and some of the solutions

---

*Presenter:*

**Bill Hall**  
*President,  
MLM Associates*

### **Track 3 - Strongly Typed Resources in Microsoft .NET**

Microsoft .NET version 2.0 and later allows the creation of strongly-typed resources for use in localization. Such resources are essentially a compiled class that contains a set of static and read-only properties. The result is an alternative to obtaining resources using methods such as the GetString or GetObject methods of a ResourceManager class. In this session, examples of strongly-typed resources along with their advantages and uses will be explained and demonstrated. Of special interest are techniques for creating and using the equivalent of satellite resources. Closely related classes will also be discussed.

---

*Moderator:*

**Deborah Anderson**  
*Researcher,  
UC Berkeley*

### **Track 4 - Unicode on the Front Lines**

This three hour session will cover a broad range of "front line" topics with short presentations and discussions. As an international character encoding standard, Unicode provides a stable format for written language documentation and interchange of text. As such, it should be the basis for projects involving written languages, to make text archivable and searchable in a standardized way. The first segment of this session will discuss three new Unicode-based projects: the effort to encode the Tai Viet script, a script used today in Vietnam, Laos, Thailand, and China, and the challenges of encoding such a script; a project to encode the historic Tangut script, used for an extinct Sino-Tibetan language of central China; and a project to develop a Unicode-based search engine for ancient Chinese text materials. All three projects demonstrate how Unicode can be used as the foundation for character encoding. The latter portion of this session will present several papers on the current use of Unicode to document endangered languages (or endangered scripts). Presentations include:

- **Challenges of Encoding the Tai Viet Script** - [Jim Brase](#), *SIL International*
- **The Tangut Encoding Project** - [Richard Cook](#), *UC Berkeley*
- **A Vast Repository for Unicode Applications over the Internet** - [Hongyuan Wang & Fred Feng Zhao](#), *Hytung Inc.*
- **Representatives from SIL (Lorna Priest), the E-MELD Project (Michael Appleby), Linguistics Dept. at the Max Planck Institute for Evolutionary Anthropology (Michael Cysouw), and the Script Encoding Initiative (Deborah Anderson)**

16:00-16:50

## SESSION 5

*Presenter:*

**Thomas Merz**  
President,  
PDFlib GmbH

**Track 1 - Unicode and PDF – Do They Play Together Well?**

It's amazingly hard to properly support Unicode in PDF! On the PDF creation side, Unicode support is mainly a matter of dealing with various font and encoding flavors. However, when it comes to extracting Unicode text from legacy documents or (even worse) from arbitrary PDFs on the Web, the job gets significantly harder. Copying Unicode text from a PDF document is useful if you want to re-purpose document contents, and is of course a crucial operation for search engines. Last, but not least, reliable Unicode text extraction is required by accessible (tagged) PDF and the international archiving standard PDF/A.

*Presenter:*

**Dale Schultz**  
Globalization Leadership  
Team,  
IBM

**Track 2 - The Social Engineering of Producing Internationalized Software**

Software does not yet write itself and perhaps it never will. Until that happens, it has to be written by humans. Getting it right is a skilled art. The technical parts are difficult enough, but there is another aspect of getting management, development and test teams to actually do what is needed. Ensuring that software is adequately internationalized is thus an exercise in social engineering. (Not the type that tries to trick people into divulging information, the type that gets people to change their behavior!) This presentation will reveal techniques that can be used during various stages of software development, from architecture through development, testing and translation.

*Presenter:*

**Alik Khavin**  
Software Design  
Engineer,  
Microsoft

**Track 3 - Internationalization Best Practices for the New Windows Presentation Foundation**

Windows Presentation Foundation, Microsoft's next-generation UI framework, offers compelling advances in UI design and content presentation making it a popular choice for developers. See real world applications built on WPF. Learn about WPF and its international features such as adaptive layout, size sharing, font fallback with composite fonts, and bidi layout features. Learn how to localize WPF applications using the platform localization APIs.

(16:00-17:50)

**Track 4 - Unicode on the Front Lines (Cont'd)**

17:00-17:50

## SESSION 6

*Presenters:*

**Matthew Hardy**  
Computer Scientist,  
**Philip Levy**  
Principal Scientist,  
Adobe Systems

**Track 1 - Unicode Issues in Mars: An XML Representation of PDF**

Mars is a new file format for PDF documents which uses XML to represent document content and metadata. The Mars file format presents a number of challenges in interpreting PDF string content and creating a valid and useful Unicode representation. PDF uses string representations in several ways: first as text content on pages, second as text in data structures that are used to describe the document structure, and third as names of objects such as fonts and images. Translating these strings into Unicode was one of the big challenges we faced in defining the Mars format.

*Presenter:*

**Edward Cherlin**  
Chairman and  
President,  
Earth Treasury

**Track 2 - Language Support on the Children's Computer**

This presentation will examine the impact of the One Laptop Per Child project on Unicode and Linux localization, present and future. OLPC is, among other things, the largest education, economic development, health, human rights, and Free Software project in the world, with a target of hundreds of millions of children and their communities. It will also be the engine for the biggest localization effort ever mounted, as OLPC XO laptops move into dozens, then hundreds and potentially thousands of language communities.

*Presenters:*

**Qianrong Ma**  
Principle MTS,  
**Makoto Tozawa**  
Principle MTS,  
Oracle

**Track 3 - Internationalization of Voice Applications**

Internationalization support for the voice application development is relatively new and its unique characteristic presents a different challenge from its GUI-base counterpart. For example, voice applications are stricter on grammar accuracy than GUI-based ones and traditional text translation process is inadequate in handling the complexity introduced by voice content. In this presentation we will discuss the special requirement on internationalization introduced by voice applications and the solution we had for the Oracle voice applications, including runtime APIs for voice application developers and build-time tools to streamline the translation and recording process for voice content.

(17:00-17:50)

**Track 4 - Unicode on the Front Lines (Cont'd)**

18:00-20:00 - IUC31 CONFERENCE RECEPTION (IN EXHIBIT AREA)

Wednesday, October 17, 2007

09:00-09:50

SESSION 7

*Presenter:*

**John Emmons**  
Senior Software  
Engineer,  
IBM

**Track 1 - What's New in CLDR 1.5**

Unicode CLDR is quickly becoming one of the most widely used and authoritative sources of localization data available to application developers and programmers. In this session, we will highlight some of the new features that have been added to the latest CLDR release. Topics include the introduction of new data to better support time zone naming conventions, relative date/time functionality, and additional supplemental data. In addition, an overview of the data submission and vetting process will be presented, along with an explanation of the current voting procedure. We will also discuss how any interested individual can become involved in the CLDR data submission and vetting process.

*Presenter:*

**Martin J. Dürst**  
Associate Professor,  
Aoyama Gakuin  
University

**Track 2 - IRIs and IDNs: Testing, Implementations, and Specification Evolvement**

Internationalized Resource Identifiers (IRIs) are the internationalized version of Web addresses. The IRI specification has been available since 2005, and the specifications for Internationalized Domain Names (IDNs) since 2003. Implementations of IRIs and IDNs in the major browsers are well advanced, but implementations for toolkits and APIs currently still leave quite a bit to be desired. This presentation will give a short general introduction to the topics of IRIs and IDNs, stressing the role of Unicode and UTF-8. It will report on an implementation effort by the author and his group for IRIs and IDNs in the widely used Web toolkit Curl, and on progress with automatic testing and automatic generation of tests for IRIs and IDNs.

*Presenter:*

**Loïc Dufresne de Virel**  
Localization Program  
Manager,  
**Michael Kuperstein**  
Localization Engineer  
**Beat Stauber**  
Localization Engineer,  
Intel Corporation

**Track 3 - Adding Unicode Support to the Intel® Viiv™ Software**

This session presents the steps taken to make the Intel® Viiv™ software fully Unicode capable. This software is a key component of the Intel® Viiv™ technology, which allows consumers to access, manage, and share their digital content across a variety of digital media devices. The session will provide an in-depth review of the technical challenges, the investigation process, and the subsequent code changes that were made by the development and localization teams in support of new features for Unicode, specific languages, and Windows Vista.

10:00 – 18:00

**Track 4 - Unicode Technical Committee Meeting**

The Unicode Technical Committee (UTC) is responsible for the development and maintenance of the Unicode Standard, including the Unicode Character Database, as well as Unicode Technical Reports and Unicode Technical Standards.

The committee meets quarterly. Since the first day of the 3rd quarter meeting this year overlaps with the conference, and is being held at the same hotel, this is a unique opportunity for conference attendees to observe and participate.

For details of the UTC, please see the UTC web page on Unicode.org [link:  
<http://www.unicode.org/consortium/utc.html> ]

10:00-10:50

SESSION 8

*Presenter:*

**Ken Lunde**  
Senior Computer  
Scientist,  
Adobe Systems

**Track 1 - Ideographic Variation Sequences: Implementation Details**

Ideograph Variation Sequences (IVSes) allow glyph distinctions to be made at the "plain text" level, through the use of the Variation Selectors (VSes) in Plane 14. This presentation thoroughly describes the implementation details for supporting IVSes in the context of OpenType fonts. In addition to the implementation details for IVSes, the experience of registering the glyphs for the Adobe-Japan1-6 ideographs will be covered during this presentation. Proper handling of IVSes and VSes, from a text-engine perspective, along with real-world application of the IVD and IVSes, are part of the overall picture that is painstakingly painted.

*Presenters:*

**Addison Phillips**  
Internationalization  
Architect,  
Yahoo!  
**Mark Davis**  
Google

**Track 2 - Language Tags: the Next Generation**

In late 2006, the IETF updated the way language tags are created and used. The new documents (RFC 4646, 4647) incorporate a number of changes to support the use of script codes, as well as a more recent update to incorporate support for ISO 639-3. This presentation, from the authors of the updated RFCs, covers the format of the new language tags and the language subtag registry; the matching algorithms for comparing language tags to user preferences; and other developments in language identification in Internet applications.

*Presenter:*

**John Brinkman**  
Software Development  
Manager,  
Adobe Systems

**Track 3 - Effective Use of the CLDR for Capturing Form Data in Adobe Systems Reader**

This session reviews the usage of CLDR data inside PDF forms. We examine three areas where Adobe has used CLDR to enhance the forms experience: 1) CLDR data/format patterns used for parsing and displaying date, numeric and currency values; 2) Allowing sections of a form to adhere to specific locales and allowing the locales to change during a session; 3) Achieving

consistent form behavior that spans operating systems, product releases and CLDR updates. The presentation will include demonstrations of Adobe Designer and Reader.

10:50-11:10 - Morning Refreshments

11:10-12:00

SESSION 9

*Presenter:*

**Tex Texin**

*Internationalization Architect,  
Yahoo!*

**Track 1 - How to be a CSI (Encoding Crime Scene Investigator)**

Join the CSI team, Grissom, Willows, Sidle, Stokes, et al. in the forensic analysis of character encoding crimes. This presentation will elaborate on the techniques of CSI forensic analysis and its application to debugging character encoding problems in software and web applications. Several example problems will be diagnosed.

*Presenter:*

**Michael McKenna**

*I18n Architect,  
Yahoo! Inc*

**Track 2 - Global Mash-ups - Dealing with Content of the World**

Humankind has been cataloging and archiving creative, scholarly, and political works since before the time of the Greeks. In the Digital Age, institutions have been storing metadata about information for the past forty years or more. Even though standards exist, and have existed for some time, each legacy repository may have chosen to store its information in different formats or encodings, may use different subsets of metadata, or different protocols to access the information. As academic and research applications bring this legacy content to light, it provides the opportunity for intriguing mash-ups of cross-cultural information on a global scale. In order to allow content access across multiple repositories physically owned and managed by different institutions, several problems must be overcome. Among these problems are normalization of metadata, font rendering, protocol recognition, cross-language queries, and mixing legacy systems with web services.

*Presenter:*

**Roy**

**Tetsuro Yokoyama**

*Principal Globalization Engineer,  
Motorola - GTG*

**Track 3 - Internationalization Programming for Mobile Applications**

In recent years, cellphone is becoming a commodity for our daily life style. Its trend is similar to what we have seen for the desktop/laptop computers where cellphones are becoming faster, providing more memory, giving the rich multi-media experiences and having a longer battery life. Interesting enough, more business professionals are realizing how capable today's smartphones have become and carry the enterprise always-connected smartphone instead of laptop. This presentation covers the overview of Unicode and locale support in various mobile platforms used in the enterprise smartphones.

12:00-13:00 - LUNCH

13:00-13:50

SESSION 10

*Presenter:*

**Murray Sargent**

*IW-Publisher/Text Services,  
Microsoft*

**Track 1 - Mathematical Input Methods**

This talk compares and demonstrates two linear-format input methods for mathematical equations along with other approaches involving handwriting, menus, toolbars and ribbons. The methods are the one used in Microsoft Office 2007, and MathTeX, a version of [La]TeX's math input with extensions and conventions for interoperating with presentation MathML and Microsoft Office's OMML. The former method favors efficient input and resembles a real mathematical notation. MathTeX favors compatibility with [La]TeX and its simpler syntax, while forgoing attempts to look like a mathematical notation. The demonstrations reveal how formula autobuildup together with WYSIWYG editing simplify and streamline equation entry.

*Presenter:*

**Craig Rublee**

*Sr. Globalization Architect,  
Adobe Systems*

**Track 2 - Creating World Ready Rich Internet Applications using Flex Builder**

The presentation will demonstrate the creation of a Rich Internet Application (RIA) using Adobe's Flex Builder that meet the requirements of an international user base and that can be easily translated into multiple languages. The use of the declarative MXML language and User Interface components that are part of the Flex Builder environment will be demonstrated as well as the use of ActionScript. ActionScript is a scripting language based on ECMAScript . An ActionScript framework that uses the CLDR for locale data will be described and demonstrated. Additionally several methodologies for creating localized versions of the Rich Internet Application will be shown. This will include facilities for run-time loading of localized resources and access to localized resources from a server.

*Presenter:*

**Katsuhiko Momoi**

*Sr. Test Egr./I18n Consultant,  
Google*

**Track 3 - Web Mail Internationalization and Unicode**

International Mail is a bewildering world with necessities to support legacy mail programs and an assortment of local encodings while adhering to Internet standards. At the same time Internet best practices and compatibility with existing mail services must be considered. There are also country and device specific requirements. Mail I18n in fact involves a number of issues that Unicode believers would rather not think about. I will present a set of requirements and suggestions that could help design your mail products better for international users and while allowing for the transition to Unicode mail in the near future.

14:00-14:50

SESSION 11

*Presenter:*

**Lee Collins**

**Track 1 - Unicode Input on Mac OS X**

Mac OS X provides a rich variety of services for inputting Unicode text. These include keyboard

*Manager-OS Engineering Asia, Apple, Inc.* layouts, character palettes and input methods for CJK and other scripts. This talk will provide a brief overview of the architecture and system support behind these services and illustrate the working of each service with demos. Topics covered will be of interest to developers considering developing text input services for the Mac platform as well as to end users who want to learn about the various options for Unicode input.

*Presenter:*

**Richard Ishida**  
*Internationalization Activity Lead, W3C*

### **Track 2 - Hints for Designing International Web Pages**

This presentation will look at a number of practical issues for people who develop web pages for a multilingual audience. Topics will include the dangers of composing sentences in content using scripting, strategies for designing layout so that text expansion during translation will not destroy your efforts, and the use of pull-down lists when navigating to localized content. We will explore some of the potential difficulties that can be encountered in these areas and pull together some best practices to help you avoid them.

*Presenter:*

**Claudia Galván**  
*Sr. Lead Program Manager, Microsoft*

### **Track 3 - Globalizing Hotmail**

Following its 10th year anniversary, Hotmail is still the largest email service on the planet with 300 million+ users covering 35 languages and 58 markets. This presentation discusses the behind the scenes globalization challenges and lessons learned in web based email systems and the route ahead.

14:50 – 15:10 - Afternoon Refreshments

**15:10 - 16:00**

**SESSION 12**

*Presenter:*

**Marc Durdin**  
*Director, Tavultesoft*

### **Track 1 - KeymanWeb - A Web-based Unicode Input Solution**

One of the emerging requirements of internationalization on the web is the provision of input solutions for foreign languages into a web page. In a mobile and multilingual society it becomes increasingly important to be able to select an input method independent of the operating system. Tavultesoft KeymanWeb implements a lightweight JavaScript multilingual keyboard solution for the web, with both custom and standards-based keyboard layouts. This presentation covers how KeymanWeb works and how KeymanWeb fits into an integrated multilingual website solution.

*Presenter:*

**Stephen Zilles**  
*Standards Guy, Adobe Systems*

### **Track 2 - Vertical Text on the World Wide Web**

This presentation will describe some of the features that enable the use of vertical text that have been or are planned for inclusion in the W3C formatting standards, such as CSS, SVG and XSL. The presenter will describe facilities for placing vertical text, aligning mixed texts, such as Western alphabetic or Arabic texts with East Asian texts.

*Presenters:*

**Shoshannah Forbes**  
*SQA Engineer, Shanjian Li*  
*Software Engineer, Google*

### **Track 3 - Gmail BiDi Enabling - A Case Study from Two Perspectives**

The Hebrew and Arabic versions of Gmail support bidirectional text editing. This talk gives a little "behind the scenes" peak into the process and issues of creating this bidi version of gmail, given from a dual perspectives- developer's and QA's. Topics include: Creating generic technical solutions for common problems (CSS changes, bracket handling, text layout order control), bug management strategies, better bug reporting enabling communication of complex issues and faster fixes and suggestions we have for future BiDi projects.

**16:10 - 17:00**

**SESSION 13**

*Presenter:*

**Michael Kaplan**  
*Technical Lead, Microsoft*

### **Track 1 - Sorting It All Out: Even More Words on Collation**

In a properly globalized product, users will have properly collated data-e.g., in the file system, in a database, in an e-mail address book. How should implementers go about ensuring culturally-correct collation in a product? What are the important linguistic issues of collation, and how do they manifest themselves in technology? This presentation goes beyond the basic tenets of collation in language, and really shows how collation functions are used (using examples from the Win32 API). It will also touch upon best (and worst) practices.

*Presenter:*

**Felix Sasaki**  
*Internationalization Activity, W3C*

### **Track 2 - Internationalization Tag Set 1.0 – A New Standard for Internationalization and Localization of XML**

This presentation introduces the "Internationalization Tag Set 1.0", a new standard recently published by W3C. We will describe how ITS 1.0 can be used to internationalize and prepare localization of XML. ITS 1.0 has a wide range of audiences: developers of new or existing XML schemas, vendors of content-related tools, and content producers. To respond to all their needs, ITS 1.0 offers powerful mechanisms that can be applied to new formats, and even to existing XML documents, without a need to modify legacy data. The presentation will provide detailed usage scenarios, applications and insights into existing and new implementations.

*Presenters:*

**Adil Allawi**  
*Technical Director, Diwan Software Limited*

### **Track 3 - Working Against the Unicode Bidi Algorithm**

This presentation discusses my experiences using and struggling against the Unicode bidi algorithm over the past few years. It will cover the practical issues that I have faced from rendering English in an Arabic newspaper to reordering similies in a mobile phone. Issues such as how to figure out the right bidi order of a stream of text that only has numbers and

symbols; How I predict if the user is thinking that a bracket on the screen is an opening or closing bracket; What to do when your check boxes are drawing in completely the wrong order and how to get a space to draw in the right place after reordering. I will also go into issues of dealing with user interface and moving a cursor through bidi text.

---

*Program is subject to change.*

- To Register for IUC31: <http://www.unicodeconference.org/registration.htm>  
Or, contact Kevin Loughry at [loughry@omg.org](mailto:loughry@omg.org)
- Exhibitor Information: <http://www.unicodeconference.org/be-exhibitor.htm>  
Or, contact Jon Roussel at [jroussel@omg.org](mailto:jroussel@omg.org)
- Sponsor Information: <http://www.unicodeconference.org/be-sponsor.htm>  
Or, Jon Roussel at [jroussel@omg.org](mailto:jroussel@omg.org), or 781-444-0404.

-