

## Program - Session Descriptions

### Monday, September 8, 2008

#### 09:00-12:30 MORNING TUTORIALS

*Presenter:*

**Richard Ishida**

*Internationalization Lead,  
W3C*

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

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

*Presenter:*

**Addison Phillips**

*Globalization Architect  
Lab126 (Amazon)*

**Track 2: Internationalization: An Introduction**

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-locales software or content.

*Presenter:*

**Shawn Steele**

**Peter Constable**

*Sr. Software Design  
Engineers  
Microsoft Corp*

**Track 3: Windows Language Support**

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.

10:30-10:45 - Morning Refreshments

12:30-13:30 - LUNCH

#### 13:30-15:30 AFTERNOON TUTORIALS

*Presenters:*

**Craig Cummings**

**Mike McKenna**

*Internationalization  
Architects,  
Yahoo! Inc.*

**Track 1 - Unicode - A Grand Tour**

This tutorial will cover the basics of what Unicode is, why it exists, and how it is used in the real world. The modules of the tutorial will cover: Introduction to glyphs, character sets, and encodings. The history behind Unicode - why was it created and what problems does it solve? The Unicode standard - what are the "Guiding Lights", or design principles behind Unicode? A tour of Unicode's structure, encoding forms, behavior, technical reports, database, and how to use the Unicode Standard. Unicode and other standards - where is it specified and why in RFCs, IETF, W3C, and elsewhere. Implementation according to Unicode - a walk through the details of attributes, compatibility, non-spacing characters, directionality, normalization, graphemes, complex scripts, surrogates, collation, regular expressions and other aspects according to the Unicode Standard and associated Technical Reports. Unicode and the Real World - an overview of International Components for Unicode (ICU) and implementations supporting Unicode in web servers, application servers, browsers, C/C++, Java, PHP, SQL, content management systems, and various operating systems. On-going programs - how Unicode is evolving to support more minority scripts, languages, and help solve linguistic processing issues.

- home
- advisory committee
- program committee
- conference program
- hotel
- registration
- press room
- become a sponsor
- IUC 32 exhibits
- become an exhibitor
- past conferences
- send me more info

**Hotel cut-off:**

**08/15/2008**

**Early-Bird  
Registration by:**

**07/31/2008**

**Venue:**

**Hilton San Jose**

300 Almaden Blvd.  
San Jose, CA 95110

*Presenter:*

**Track 2 - Web Internationalization - Standards and Best Practices**

**Tex Texin**

*Technical Director,  
NetApp*

**(13:30 to 16:30)**

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:*

**Track 3 - Internationalization with Java and Eclipse**

**George Rhoten**

*Software Developer  
IBM*

Java, Eclipse and ICU have excellent frameworks and tools to simplify your job of internationalizing software. The Externalize Strings Wizard in Eclipse provides an easy to use interface for extracting strings from source code into your resource bundles. This tutorial will discuss this wizard, resource bundle management, formatting messages and how the default locale can changes resource lookup and Java framework behavior.

15:30-15:45 - Afternoon Refreshments

**15:45-17:45**

**AFTERNOON TUTORIALS**

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

*Presenter:*

**Track 2 - Creating XHTML/HTML Pages with Right-to-Left Scripts**

**Richard Ishida**

*Internationalization Lead,  
W3C*

**(16:45 to 17:45)**

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

*Presenters:*

**Track 3 - ICU in Action**

**David Bertoni**

*Software Engineer,  
Google*

**Steven R. Loomis**

*Software Engineer,  
IBM*

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.

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

**Tuesday, September 9, 2008**

**09:00-09:15**

**WELCOME & OPENING REMARKS**

*Addison Phillips, Globalization Architect, Lab126 (Amazon)*

**09:15-10:00**

**KEYNOTE Presentation: New Developments in Digital Humanities**

*Georges Van Den Abbeele, Dean of Humanities University of California, Santa Cruz*

**10:00-20:00 - EXHIBIT AREA OPEN**

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

**10:30-11:20**

**SESSION 1**

*Presenter:*

**Track 1 - Unicode as a Key Tool in Preserving Archimedes Writings**

**Doug Emery**

*Consultant  
Emery IT*

**Michael B. Toth**

*Program Manager,  
R.B.Toth Associates*

Unicode has proven to be a key tool for encoding transcriptions of the earliest known texts of Archimedes' key mathematical and scientific works. As part of a major multiyear project using a range of advanced imaging techniques, the transcriptions are being encoded in Unicode to make them readily available on the Web for users around the globe. Integrating transcriptions of Archimedes' mathematical texts with multi-spectral digital images and hosting them on the Web for global users has posed a complex set of information sharing challenges. Unicode is an essential element in the integration of the transcribed information with the archive of complex digital images.

*Presenter:*

**Addison Phillips**  
Globalization Architect  
Lab126 (Amazon)

### Track 2 - WS-I18N: Making Web Services Internationalized

Web services and internationalization have an uneasy relationship. Whether you use REST, AJAX, or SOAP, it isn't always clear how to extend your internationalized code to "live at the end of a URI". This presentation details both the latest developments on the WS-I18N at the W3C as well as some ideas on how to develop (and examples thereof) REST/AJAX Web services.

#### Presenters:

**Jeffrey D. Oldham**  
Software Engineer  
**Dr. Craig Cornelius**  
International  
Engineering Team  
Google, Inc.

### Track 3 - Dealing with a World in Flux: Updating International Identifiers

Google uses identifiers to denote a user's language, region (a.k.a., country), currency, and time zone. Because sets of valid identifiers frequently change, Google incorporates these as quickly as possible while still supporting deprecated identifiers. We describe our engineering to support updating identifiers. Region identifier transitions are the most difficult to implement, especially in the Ads system which uses many interacting executables. Based on a taxonomy of identifier changes, we engineer this process using a human-executed, distributed transition plan and a small number of code updates. Users of identifiers may benefit from Google's experience.

11:30-12:20

SESSION 2

#### Presenters:

**Swaran Lata**  
**Somnath Chandra**  
Scientist-D  
Dept. of Information  
Technology,  
Ministry of  
Communications &  
Information Technology  
Gov't. of India

### Track 1 - Challenges of Localization in a Multi-script and Multilingual Nation

The multilingual diversity of India having twenty-two officially recognized languages and 11 scripts is probably the most unique in world making internationalization and localization of any software solution a complex and gigantic task. We shall discuss Storage and Encoding problems, then input mechanism and storage mapping in respect of Indic languages, especially non-existence of a unique and converged storage location from different keyboard layouts. The rendering and browser problems specific to Indic languages would be presented. The requirements of Indic languages in respect of IDN, special modifier characters, and limitation of string-length in the light of latest IDNA200x protocol will be presented. Finally we will touch upon various other challenges in Localization such as Orthographic variation, long gestation period for convergence and skewed participation of stake holders in adopting and implementing standards.

#### Presenters:

**Adil Allawi**  
Technical Director,  
Diwan Software Ltd  
**Shanjian Li**  
SW Engineer  
Google

### Track 2 - I18n in Google Web Toolkit - An Open Source Collaboration

Google Web Toolkit (GWT) is an open source Java software development framework that makes writing AJAX applications simple. So says the first line of the the GWT home page. But many companies slap on an open source label on their projects without really meaning it. When one developer came along with the aim of having Arabic supported in GWT he found Google truly open to his ideas and this presentation demonstrates the results of that collaboration. Adil Allawi and Shanjian Li will discuss the new I18N features of GWT 1.5.

#### Presenters:

**Addison Phillips**  
Globalization Architect,  
Lab126 (Amazon)  
**Dr. Mark Davis**  
Sr. Intl. SW Architect,  
Google

### Track 3 - Language Tags: The Next Generation

In 2006, the IETF issued an updated version of BCP 47 "Tags for Identifying Languages", which updated the way languages are identified in most computer programs and protocols. BCP 47 now incorporates major changes to language identification, including many more base languages, the use of scripts to distinguish written forms, the addition of more regional variations, and the ability to effectively parse language identifiers. This presentation, from the authors of the updated RFCs, covers the format of language tags and the language subtag registry; the matching algorithms for comparing language tags to user preferences; the new features in BCP 47 and their impact on developers; and other developments in language identification in Internet applications.

12:30-13:30 - LUNCH

13:30-14:20

SESSION 3

#### Presenter:

**Dr. Mark Davis**  
Sr. Intl. SW Architect  
**Andy Heninger**  
Software Engineer  
Google

### Track 1 - 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 updated 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:

**Jim DeLaHunt**  
Consultant in World-ready  
Business and Technology  
Development  
Jim DeLaHunt &  
Associates

### Track 2 - Web 2.0 Goes to Babel: Multilingual Websites and User-supplied Content

In today's web, it's straightforward to publish in any single language. The cool Web 2.0 sites are organized around user-supplied content: postings, tags, comments, photos, videos. But what happens when you try to do all that in more than one language at a time? Do you translate the user-supplied content? And how? Can you crowdsource the localization? This talk looks at the business, technical, and design issues of multilingual web sites. We'll look at role models, examine social translation, see how technologies like Joomla, Drupal, WordPress, HTTP, and URLs fit in. Get inspired to add another language to your site!

---

**Presenter:**

**K.G.Sulochana**

*Joint Director*

**Kumar R. Ravindra**

*Senior Director*

*Language Technology Ctr.*

*C-DAC,*

*Thiruvananthapuram*

### **Track 3 - IDN for Indian Languages - A Case Study**

India has perhaps the richest mix of living languages with 4 major language families, 22 official languages, many unofficial languages and more than 2000 dialects. With less than 10% English literates among her population, the Internet penetration in India is around 5% only. IDNs and local language web content can play an important role in increasing the Internet usage. But the complex nature of Indic scripts/languages pose problems in IDN implementation. This presentation will discuss IDN for Indian languages in general and the problems likely to be encountered in the implementation, particularly the security issues, with examples from South Indian languages.

---

**14:30-15:20**

**SESSION 4**

**Presenter:**

**Addison Phillips**

*Globalization Architect*

*Lab126 (Amazon)*

### **Track 1 - What's in a Name? Handling Personal Names and Information in a Global Application**

People's names, their presentation, collection, collation, and validation, are rich in cultural and linguistic variation and nuance. Handling people's personal information (which may also include gender, age, and other related information, as well as regulatory concerns) is a key problem when internationalizing an application that deals with this type of information. This presentation gives an introduction to the variations in name handling and demonstrates some different approaches to designing multi-lingual, multi-culturally capable systems.

---

**Presenter:**

**Katsuhiko Momoi**

*Senior Test Engineer*

*& I18n Consultant*

*Google, Inc.*

### **Track 2 - Web Apps I18n Testing and Test Data**

Quality of data has a major impact on the effectiveness of i18n testing - manual or automated. In this talk, against the background of web apps development, I present a comprehensive list of data types needed for i18n testing, how they should be used, and how to generate them. I discuss in particular what additional considerations are needed for i18n data use in current web apps testing environments - rapid development cycles, frequent code changes, effective testing for a large and expanding number of localized languages, easy use for automated testing, compliance for the latest Unicode standard, etc.

---

**Presenters:**

**Richard Ishida**

**Steven Zilles**

**Tatsuo Kobayashi**

*W3C & Adobe*

### **Track 3 - New Work on Japanese Layout Requirements**

The W3C has been gathering requirements for Japanese text layout from in-country experts, including the original authors of JIS X 4051 via its Japanese Layout Task Force. The task force also includes representatives from the CSS, XSL-FO and SVG Working Groups. The end result will be a detailed set of requirements that will build on and extend beyond JIS X 4051, and will be published as a W3C Note in English and Japanese.

The work has revealed some interesting new insights on Japanese layout for the Western experts involved. This talk will describe these and some of the key points of the proposed document, referring to rules for such things as page set up using kihan hanmen, the amount of blank space associated with punctuation and parentheses, text justification and line breaking, conversion between horizontal and vertical text, etc.

---

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

**16:00-16:50**

**SESSION 5**

**Presenter:**

**Loïc Dufresne de Virel**

*Localization Strategist*

**Michael Manca**

*Localization QA Lead*

**Cory Whitney**

*Localization Engineer*

*Intel Corporation*

### **Track 1 - We're "World-Ready"... What does this really mean?**

Proper internationalization is routinely listed in most software requirement documents, but most development and validation teams are in the dark when the time comes to implementing and testing this basic requirement. Based on real software bugs investigated by Intel's localization team, this session presents the typical internationalization issues that developers encounter every day, but often struggle to properly address in a proactive fashion, prior to an actual localization attempt. Regional settings, language selection, encodings, and UI design are among the topics that will go under the microscope in a very practical way, exploring the probable causes of those issues, along with possible solutions and best known methods implemented at Intel.

---

**Moderator:**

**John C. Emmons**

*Globalization Architect*

*IBM Software Group*

### **Track 2 - Panel - Unicode Locale Data**

The Unicode Common Locale Data Repository (CLDR) is by far the largest and most extensive standard repository of locale data, used by a wide spectrum of companies for internationalization and localization of applications and systems. This session will discuss what types of locale information are available in CLDR, including the new data available in CLDR 1.6; the LDML language specifying the data; how the data are intended to be used; how the CLDR vetting process works to ensure the quality of data; and how interested individuals can become involved in the project. Panelists will also discuss how they are making use of CLDR data, and discuss issues in the collection and production of data.

The panel will consist of persons from multiple vendors involved in deploying CLDR in their own products and projects, as well as those involved in the data gathering and vetting

process. Comments and questions will be welcomed from the audience.

Panelists: TBA

**Presenter:**

**Ken Lunde**

*Sr. Computer Scientist  
Adobe Systems*

**Track 3 - Ideographic Variation Sequences: Implementation Details & Demo**

Ideographic 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 'cmap' tables of OpenType fonts. The ideographs (aka, kanji) in the Adobe-Japan1-6 character collection, a static glyph set that was set forth by Adobe Systems, represent the very first set of glyphs to have successfully gone through the Ideographic Variation Database (IVD) registration process, and whose IVSes became registered at the end of 2007. In addition to covering the implementation details for IVSes, the presentation spends a significant amount of time conducting a live demo of IVS-enabled OpenType Japanese fonts being used with a variety of IVS-savvy applications.

**17:00-17:50**

**SESSION 6**

**Presenter:**

**Roy Yokoyama**

*Principal Globalization  
Engineer  
Motorola-GTG*

**Track 1 - Internationalization Programming for Mobile Applications**

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

**Track 2 - Panel - Unicode Locale Data (Cont'd.)**

**Presenters:**

**Ken Lunde**

*Sr. Computer Scientist  
Adobe Systems*

**Track 3 - Legacy Gaiji Solutions & SING**

The so-called gaiji problem lingers, yet there have been numerous gaiji solutions over the years. Adobe Systems carefully and painstakingly studied legacy gaiji solutions, noted their shortcomings, and gave birth to SING, an acronym for "Smart INdependent Glyphlets." This presentation details the shortcomings of legacy gaiji solutions, pinpointing the specific areas that can cause them to fail in open systems and environments, and provides details about how SING has become a very effective gaiji solution for open systems."

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

**20:00-21:00**

**"Birds of a Feather" Discussion**

**Moderators:**

**Erkki I. Kolehmainen**

*Independent Consultant,  
Cultural Diversity Issues  
in ICT*

**Tero Aalto**

*Coordinator, Kotoistus  
Initiative  
CSC-The Finnish IT Center  
for Science*

**Soft Cultural Elements in ICT**

CLDR collects culturally dependent data. Its data structure is defined by LDML, and the data is available for use in IT implementations. A European CEN Workshop has been tasked to define sharing of e-governmental resources. For this, it plans to define (in co-operation with CLDRTC and LISA) a registry to cover "Soft Cultural Elements" with global relevance, such as: - mapping of countries/regions and languages, and their usage and proficiency; - forms and usage of written and spoken salutations; - rules relating to personal names, form and usage; - holidays and their coverage, including whether offices/shops may be expected to be closed; - rules relating to written text, such as highlighting conventions, etc.; - timeliness expectations; - any culture-specific requirements relating to form, appearance, color, etc. This BOF deals with these elements and some such structure that could meet the cultural and linguistic requirements of both human and machine users.

**Wednesday, September 10, 2008**

**09:00-09:50**

**SESSION 7**

**Presenter:**

**Michael Ow**

*Software Engineer  
IBM*

**Eric Mader**

*Sr. Software Engineer*

**Track 1 -**

Matching, sorting, and searching Unicode text presents a unique set of challenges for software internationalization. Different languages, even those written with the same script, have different rules for how strings should be compared. This talk will describe the different rules used to compare strings in different languages and give a brief description of the Unicode Collation Algorithm showing how it can be used to implement these rules. We will discuss the efficient searching of Unicode text examining some of the more popular string

**Presenter:****Kirti Velankar**

Senior Software Engineer  
Yahoo Inc.

**Track 2 - Internationalization Support in PHP**

PHP is one of the most popular platforms for modern Web development. This session gives an overview of the internationalization support in PHP 5 and PHP 6 as well as describes new developments and future plans in this area.

**Presenters:****Dr. Deborah Anderson**

Researcher  
Dept. of Linguistics;  
Project Leader  
Script Encoding Initiative  
UC Berkeley

**Don Osborn**

Director, Bisharat

**Track 3 - Unicode Support for Modern African Languages and Scripts: Where are We Today?**

Africa is home to a large number of languages, with the figure of over 2000 often cited. Due to economic and technological hurdles--as well as the sheer number of languages--being able to provide widespread software and font support for the orthographies of African languages with Unicode is challenging. This talk will provide an overview of the present-day situation, and focus on the current needs in terms of fonts, display, and input, as well as information on unencoded scripts.

10:00-10:50

SESSION 8

**Presenter:****Yoshito Umaoka**

Software Engineer  
IBM

**Track 1 - Knotty Problems in Date/time Parsing and Formatting and Time Zones**

An internationalized UI must be able to format dates and times in a localized form according to different languages and regional conventions. Programs need not only the choice of short or long formats, but also different choices of fields, such as "December 13, 2008" vs "Dec 2008" or "12/13", and a way to format ranges of dates ("Dec. 12-14"). This presentation will explore problems in date and timezone formatting, and discuss how they are addressed by structure and data in the Unicode Locales project. It will provide recommendations for implementation and usage in various practical scenarios.

**Presenter:****Douglas R. Davidson**

Software Engineer  
Apple, Inc.

**Track 2 - Apple's Architecture for Localization and Internationalization**

Apple's operating systems are designed with top-to-bottom international and multilingual support. This session covers the underlying foundations of that support, including Unicode strings, localization, locale data, collation, fonts, and text display, from both a developer and a user perspective. Demonstrations will be given of the capabilities of Mac OS X's latest version, Mac OS X 10.5 Leopard. For developers, examples will be shown of the use of many of the relevant application programming interfaces.

**Presenter:****Dr. Richard Cook**

Linguist  
UC Berkeley

**Track 3 - Unencoded Scripts of China**

This presentation introduces scripts of China that are not yet supported by Unicode. These are of several different types ("ideographic", pictographic, syllabographic, alphabetic), and have progressed through various stages of the standardization process (exploratory research, draft proposal, final proposal). In particular, we will examine details of Tangut (Xixia), Nüshu, Naxi, Lisu, and Classical Yi. Each of these writing systems presents Unicode with unique challenges. Some character sets are very large, some have surface similarity to characters already encoded, and some are unlike anything previously digitized.

10:50-11:10 - Morning Refreshments

11:10-12:00

SESSION 9

**Presenters:****Martin J. Dürst**

Associate Professor

**Kunihiro Sato**

Aoyama Gakuin University

**Track 1 - Implementing Better Source Editing for Bi-directional XHTML and XML**

This presentation describes an ongoing implementation effort to make it possible for the first time to edit the source of bi-directional XHTML and XML documents. The implementation is based on an approach developed using a Web-based prototype. The ease of viewing and editing (X)HTML and XML source was one of the main reasons for the fast adoption of these technologies. However, for scripts and languages written right-to-left, such as Arabic and Hebrew, very serious obstacles for source editing have remained. The root of the problem is that syntax-significant characters, such as angle brackets and quotes, are weak or neutral, which may lead to very confusing display situations. Our implementation uses syntax-based context analysis to change the bi-directional type of some weak characters as part of a higher level protocol.

**Presenter:****Michael Bridgers**

Senior Software Engineer  
SAS Institute Inc.

**Track 2 - Shikari: Hunting for Java I18N Problems**

SAS Institute's Shikari tool is used to search for i18n problems in Java applications. (Shikari is a Hindi word that means Hunter.) This presentation will show what types of i18n problems Shikari detects, how SAS uses Shikari in its development process, and how the Shikari tool is organized and extended. Shikari does static analysis of Java applications to identify and correct i18n problems as the code is being developed. Shikari is built as Eclipse plugins, and runs in two modes: 1) As plugins to the Eclipse Java IDE. 2) As a command line Eclipse RCP application.

**Presenter:****Michael Kaplan**

Software Development  
Engineer  
Microsoft Corp.

**Track 3 - Behind the Proposed Change to Tamil in Unicode**

The encoding of Tamil within Unicode has been the subject of displeasure by the government of Tamil Nadu for as long as it has been there. It has led to a proposal (built up over the last decade) to try to change the way that Unicode looks at Tamil, and the very real questions of why this effort has been so persistent and what will eventually happen have not really been discussed overtly in all of this time. This presentation's goal is to talk

about why the proposal exists, why it will ultimately fail, and why the language itself can survive that fact. The broader issues of the view of languages and the "rights" of language owners will also be discussed in this case study of a language that has been both wronged and righted as few others have in modern times.

12:00-13:00 - LUNCH

13:00-13:50

SESSION 10

*Presenters:*

**Shawn Steele**

*Sr. Software Design  
Engineer*

**Poornima Priyadarshini**

*Program Manager  
Windows International  
Microsoft*

**Track 1 - Globalization & Microsoft Silverlight**

In this presentation we will discuss the core globalization elements of the Microsoft Silverlight platform, including the dependency on the host operating system. We will demonstrate the variation of the Silverlight globalization behavior between the Microsoft Windows and Apple OS X environments. This will be contrasted to the consistent support in Microsoft .Net and Windows which each carry their own data set. The specific globalization aspects being covered will include the impact of differing culture sets, sorting behavior and character support.

*Presenter:*

**Tex Texin**

*Technical Director  
NetApp*

**Track 2 - Honey, My Unicode Data Disk Went into the Circular File!**

This session will present some of the difficulties of providing a common international interface to file services on different operating systems. Although Unicode supports all the necessary characters, identifying the set of characters that are legitimate on any OS can be difficult, and rules for case-insensitivity, normalization, etc. vary, and may even vary by user. At this time, the conclusion is unclear. The presentation may just describe the problem space, it may also offer some solutions, and possibly a proposal for standardizing filename conventions will be offered.

*Presenter:*

**Wunna Ko Ko**

*Project Leader  
Burmese Language  
Projects*

**Track 3 - Unicode: A Ray of Hope For Myanmar Scripts Community**

There is no localized international software in Burmese language which is used by 50 million people. Myanmar script represents not only Burmese script but also Mon, Shan, Karen scripts which share a lot of commonalities. Unicode 5.1 is supposed to be a complete encoding for Burmese script but a complete implementation method is not yet available. Besides the encoding standard, more work needs to be done. This presentation will discuss who is going to do that work and what more should be done to increase support of the Myanmar script.

14:00-14:50

SESSION 11

*Presenter:*

**Owen Yen**

*International Program  
Manager  
Microsoft*

**Track 1 - Windows Live Messenger Internationalization**

Did you know that Windows Live Messenger supports 48 languages and over 100 markets, and that over 90% of the instant messaging traffic is being done by users outside of the United States? How does Windows Live Messenger behave differently according to various factors such as language, market, etc.? Come and learn about best practices used in the globalization and international deployment of Instant Messaging in general, with a case study focusing on Microsoft's Windows Live Messenger.

*Presenter:*

**Gisle Forseth**

*SW Systems Developer  
Sr. Principal  
ACUCORP  
a Micro Focus Company*

**Track 2 - Unicode and ISO 2002 COBOL, the Meeting of Two Standards**

As one of the oldest programming languages, COBOL has never had a standardized way of handling extensive character sets such as the Unicode standards. In 2002 the ISO standard organization released the ISO 2002 COBOL standard, in which the language syntax had been extended to include support for Unicode and locales. This is the story of an implementation of the modern character encoding standard into a programming language as old as the computer. Topics covered include use of the ICU libraries, supporting multiple encodings, portability, matching theory with real world and stepping waters, "where no man has been before".

*Presenter:*

**Dr. Seyed Mohamed  
Buhari**

*Senior Lecturer  
Dept. of Computer Science  
University Brunei  
Darussalam*

**Track 3 - Arwi: Case Study of Arabic, Syriac and Diacritical Unicode Characters**

This presentation will start with addressing the motivation behind the need for Arwi Unicode development. Keeping the motivation and history of Arwi script in mind, the audience will be made aware of its difference from Arabic language script. Combining Arabic, Syriac and Diacritical characters to build an Arwi script and its issues faced on different operating systems like Windows XP/Vista and Linux variants will be discussed. Rendering issues related to Arwi script on different editor software will also be addressed. Presentation will address the development of Arwi TrueType font using Fontforge software and the development of keymap for various operating systems.

14:50 – 15:10 - Afternoon Refreshments

15:10 - 16:00

SESSION 12

*Presenters:*

**Chris Weber**

*Security Consultant  
Casaba Security*

**Track 1 - Exploiting Unicode-enabled Software**

This talk will showcase some of the ways that Unicode has been leveraged to cause software to break. We will survey the security issues outlined in Unicode Technical reports 36 and 39. The issues highlighted will be illustrated by examples of historical Unicode-related security flaws in popular software and Web applications. For each vulnerability we will assess the damage that was inflicted, describe how the exploit worked, and discuss the root cause. Examples will include demonstrations of how clever attackers can exploit Unicode-enabled software to run arbitrary code or takeover the machine.

*Presenter:*

**John Harvey**  
Software Engineer  
Apple Computer Inc.

### **Track 2 - Building Input Methods on Mac OS 10.5 and Up**

Leopard introduced the Input Method Kit, a framework designed to make it easier for developer's to build input methods. This session will take you through the process of creating an input method. A basic overview of the Input Method Kit will be presented. Additionally, a complete input method will be created and built using the Input Method Kit and Apple's development tools. The new support for plug-in Input Methods will be covered. Finally, a simple plug-in input method will be created.

*Presenter:*

**Shaloo Chaudhary**  
Program Manager  
**Hisami Scott**  
International Test Engineer  
Microsoft Corporation

### **Track 3 - Internationalization of Web Applications Using Natural Language**

Have you talked to software chat-bot yet? Today chat-bots can speak multiple languages using Natural Language Processing (NLP). Natural Language Processing is the latest trend for developing Web Applications. This talk walks through best practices to develop and support international chat-bots using NLP, demonstrations will use real life example of developing a Windows Live Agent (chat-bots). We will also cover the challenges in scaling a web application using NLP to multiple languages quickly.

**16:10 - 17:00**

**SESSION 13**

*Presenter:*

**Andy Heninger**  
Software Engineer  
Google

### **Track 1 - Unicode Security and Spoofing Detection**

Because Unicode contains such a large number of characters and incorporates the varied writing systems of the world, incorrect usage can expose programs or systems to possible security attacks. This is especially important as more and more products are internationalized. This presentation describes some of the security considerations that programmers, system analysts, standards developers, and users should take into account, and discusses approaches to reduce the risk of problems. Particular attention will be given to techniques for detecting spoofing, where URLs or other identifiers may not be what they visually appear to be because of the use of characters that can be readily mistaken as others.

*Presenter:*

**Steven R. Loomis**  
Software Engineer  
IBM

### **Track 2 - What's New with ICU?**

This presentation will provide a brief overview of the International Components for Unicode library. ICU is the premier Unicode-enablement software library, providing a full range of services for supporting internationalization consistently across multiple platforms, with C, C++, and Java APIs, as well as being freely available as open source. This presentation will provide a brief overview of ICU, with emphasis on the new ICU 4.0 release, which includes the latest support for Unicode 5.1 and CLDR 1.6.

*Presenter:*

**Daniel Yacob**  
Semantic Solutions  
Architect  
TopQuadrant  
(Ge'ez Frontier Foundation)

### **Track 3 - S13N - Enabling Unicode Standards for the Semantic Web**

This talk will present an initial attempt at porting Unicode knowledge (standards, specifications, reports, and data sets) along with ISO dependent standards, to the standards of the Semantic Web as a family of ontology models. In a phrase - the "Semanticization" (S13N) of Unicode. Interoperability with ontologies of various domains and disciplines will be demonstrated followed by discussion of new capabilities that emerge with Unicode as a knowledge base. A particular focus will be on the relevance of linked data on L10N. Some background knowledge on the Semantic Web is assumed for the talk.

*Program is subject to change.*

- To Register for IUC32: <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 Ken Berk at [ken.berk@omg.org](mailto:ken.berk@omg.org)
- Sponsor Information: <http://www.unicodeconference.org/be-sponsor.htm>  
Or, Ken Berk at [ken.berk@omg.org](mailto:ken.berk@omg.org), or 781-444-0404.

Object Management Group, (OMG) organizes the Internationalization and Unicode Conferences around the world under an exclusive license granted by the Unicode Consortium. Personal information provided to OMG via this website is subject to OMG's [Privacy Policy](#). All responsibility for conference finances and operations is borne by OMG. The independent conference board provides technical review of the program and papers. All inquiries regarding the Internationalization and Unicode Conferences should be addressed to [info@unicodeconference.org](mailto:info@unicodeconference.org). Copyright © 2008 Object Management Group. All Rights Reserved.