Hotel cut-off :
October 27, 2006

Venue:

Hilton Washington
Dulles Airport

Send me more
information
on
IUC 30

Program

Wednesday, November 15, 2006

09:00-10:30	MORNING TUTORIALS

Presenter:

Asmus Freytag
President
ASMUS, Inc.

(09:00-09:45)

Track 1: Unicode 5.0 Tutorial: Part 1 - Characters in Action
Part I of the Unicode 5.0 Tutorial is a uniquely accessible and entertaining way of visualizing the core concepts of the Unicode standard. In this part you will find answers to these questions: What is a Unicode character and how are Unicode characters represented and used in a modern computing environment? How are Unicode characters entered into and displayed on a computer? How are Unicode characters interchanged? What is the interaction between Unicode and rich text (markup)? How do end-users experience Unicode? Throughout Part I, the Unicode 5.0 Tutorial gives typical examples of how the Unicode Standard interacts with the other elements of an internationalized software architecture. With the help of concrete scenarios for the use of Unicode characters you will become familiar with the role the Unicode Standard plays and the benefits of supporting it. Part I of the tutorial provides a concrete context to which the more systematic and detailed treatment of the features of the Unicode Standard presented in Part II and Part III can be related.

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:

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.

10:30-10:45	Morning Refreshments

10:45 – 12:15	MORNING TUTORIALS (Cont'd.)

Presenter:

Richard Ishida
*Internationalization Activity Lead
W3C*

(09:45-12:15)

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.

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

Track 3 - Web Internationalization - Standards and Best Practices (Cont'd)

12:15-13:15	LUNCH

| 13:15-15:30 | AFTERNOON TUTORIALS |

Presenter:

Asmus Freytag
President
ASMUS, Inc.

Track 1 - Unicode 5.0 Tutorial: Part 2 - Fundamental Specifications

Part II of the Unicode 5.0 Tutorial builds on the concepts introduced in Part I and 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. Part II 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.

Presenters:

Elsebeth Flarup
Globalization Architect
IBM Corp.

Soeren Bendtsen
Advisory IT Specialist
IBM Corp.

Track 2 - Best Practices in Software Localization Process and Technology

Software localization is often treated as a separate process which only starts towards the end of the development cycle, and in many companies the localization group does not function as an integral part of the development organization. This tutorial uses practical examples to demonstrate how localization that is built into the process, starting from the design phase, may help lower cost and improve time-to-market for localized versions.

Topics include: Building international support into the product from the beginning; Globalization and localization considerations when selecting programming platforms, tools and third-party applications; Globalization verification testing and how to use pseudo localization effectively; Do's and don'ts when creating translatable text; Translation file formats – recommendations and handling of non-standard file formats; Translation file check tools and how they can reduce translation problems, build issues and test duration; Translation planning systems, workflow and problem reporting; Translation verification testing; Source control and change freezes; Terminology management; Computer aided translation tools such as translation memory based systems. Demos will be used to illustrate tools and processes.

Presenter:

Bill Hall
MLM Associates, Inc.

Track 3 - Internationalization and Localization Features of Microsoft .NET

The internationalization and localization paradigm of Microsoft .NET is quite different from the older Win32 model, mandating a change of viewpoint and a new set of skills for programmers, localizers, and architects. In Windows you depend on a loose collection of API's and manifests to take care of globalization issues and managing resources. Now your tools are globalization and resource classes and related enumerations and the system is entirely Unicode based. If you are considering

a project in .NET, take the time to learn the basics about how the globalization features are organized and managed. You will realize that .NET is quite well designed, and practically any developer can learn the subject well enough to do the right thing most of the time. Indeed, thanks to the .NET model, most everything happens automatically and you can rely on the default behavior most of the time. But, if needed, you will also learn to recognize when an override is necessary. You may not emerge as an expert, but you will know the basics. You will be especially interested in the features of .NET 2.0 including the ability to modify locales and applying strongly typed resources.

15:30-15:45	Afternoon Refreshments

15:45-18:00	**AFTERNOON TUTORIALS**

Presenter:

Asmus Freytag
President
ASMUS, Inc.

Track 1 - Unicode 5.0 Tutorial: Part 3 - 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. Part III 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. Part III 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. Part III 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:

Eva Sanchez
Director of Project
Management
Basis Technology

Track 2 - Managing Internationalization Projects
This tutorial describes the life cycle of an internationalization development project. It will commence with the Planning phase, followed by the Execution Phase and finish with the Delivery Phase and preparation for future releases. The goal of this tutorial is to prepare the project management team as well as engineering and QA management teams to anticipate the specific challenges of an internationalization project and acquire mitigation strategies to ultimately manage to completion a successful project. The tutorial will not focus in one specific case study but the presenter will bring real life examples into the discussion. The intended audience is project managers, engineering

managers, QA managers and product managers. The level is Beginner/Intermediate

---

Presenter:

Pierre Cadieux
President
i18N Inc.

Track 3 - 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.

Thursday, November 16, 2006

09:00-09:15	***WELCOME & OPENING REMARKS***
	Mark Davis - **President, Unicode Consortium**

09:15-10:00	**KEYNOTE - One Laptop per Child — $100 Laptop** **Nicholas Negroponte** **Chairman, One Laptop per Child and Chairman Emeritus, MIT Media Laboratory**

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

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

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

Presenter:

Richard Ishida
Internationalization Activity Lead
W3C

Track 1 - Internationalizing Style on the Web: What's Planned for CSS3
Work on CSS3 is getting nearer to completion and will provide a huge leap forward for rendering and styling non-Western text. Many of the innovations are focused specifically on the needs of Asia and include such things as vertical text, grid layout, auto spacing, sophisticated white-space handling and text wrapping, justification, advanced list styling, styling for ruby, emphasis, kumimoji, and more. This presentation will show you examples of a number of these developments, and discuss how we need help to move the work forward.

Presenter:

Track 2 - Undetecting the Encoding: Choosing a

Addison Phillips
Internationalization Architect
Yahoo!

Legacy Encoding Algorithmically

Detecting the character encoding used by a block of text is an important first step in being able to process or display the text correctly. Frequently, applications will use the encoding to convert to Unicode in order to enable multi-lingual support. But what happens when you want or need to convert a Unicode text stream to a legacy encoding? Which encoding should you use? How can you reliably detect the "best" or most widely supported encoding for a given text buffer? This presentation covers a method for detecting the best legacy encoding to use when converting a UTF-8 text buffer. It describes the encoding choices to be made; the reasons one might want to do this; the interaction between "superset" and "subset" encodings; and the data structures used to get peak performance with the least amount of memory.

Presenter:

Gwyneth Marshall
International Program Manager, Office
Microsoft Corporation

[Top]

Track 3 - Microsoft Office 2007 – International Collaboration and Analysis

Building upon the strong support that Office already gives for worldwide documentation standards, this new version of Office offers an even better user experience with deeper international functionality. In this presentation we will discuss newly supported locales as well as the depth of locale support across Office documents. Office 12 offers a language neutral architecture for improved multilingual support, searching, and deployment, which will be discussed in the presentation. Of particular interest will be the new file formats that support easier data entry, retrieval, and indexing. In addition, we will give an overview of some of the new international features in Office 2007, including the Language Reference ToolTip and the English Writing Assistant. We continue to expand our work with Language Interface Packs, and will highlight the new UI languages that will be supported in Office 2007.

11:30-12:20   SESSION 2

Presenter:

Raghuram Viswanadha
Software Engineer
ICU Team
IBM Corp.

[Top]

Track 1 - StringPrep: Unicode in Network Protocols

Network protocols require consistent comparison of strings. The StringPrep
framework (RFC 3454) facilitates this function. It provides sets of rules
that can be applied to strings to prepare them for use in any protocol or
program. Each system sets up a profile of StringPrep by selecting a set of
rules. Important profiles such as NamePrep, NFS, ResourcePrep, NodePrep
are explained. The usage of StringPrep and IDNA frameworks is illustrated
by implementation in International Components for Unicode (ICU).

Presenter:

Richard Gillam

Track 2 - Parsing Personal Names

Applications that do things with people's names usually need to parse them.

*Software Engineer
IBM Corp.*

For example, lists of names are usually sorted by surname, so you need to
identify the surname. You may also need to remove or otherwise handle
extraneous components, such as titles, that often travel with a name but
aren't technically part of it. You may also need to disentangle records
that combine more than one name in a single string (such as "John and Mary
Smith"), or identify names with missing components ("John"). And in a
global environment, you need to be able to do these things correctly with
names from a wide variety of international cultures, each with its own
conventions for naming people. You may also be faced with a database of
names that are already divided into given name and surname, but
inconsistently or incorrectly. This talk will discuss all the
various things that make parsing personal names
complicated, taking an in-depth technical look at
NameParser, a component of IBM's Global Name
Management suite, and the techniques it uses for making
sense out of personal names from different cultures.

Presenter:

Michael Kaplan
*Technical Lead
Microsoft
Corporation*

Track 3 - Keyboards on Win32: Beyond the Keyboard Layout Creator
Creating Microsoft Keyboard Layout Creator (MSKLC) was a great step forward for ease of developing keyboards on Windows. But there are plenty of features and problems (and bugs!) that are not covered by this tool, some of which are not really addressable even in future versions as they are core to Windows itself. These issues can haunt both users and developers alike unless they are understood, which is where this presentation comes in -- a review of additional problems and their solutions to make the keyboarding experience on Windows better for both users and developers.

12:30-13:00   **LUNCH**

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

Presenter:

Steven R. Loomis
*Globalization
Center of
Competency, IBM
Corp*

Track 1 - Collation in ICU
Collation is the general term for the process of determining the sorting order of strings of characters for a given language. It is a key function in computer systems; whenever a list of strings is presented to users, they are likely to want it in a sorted order so that they can easily and reliably find individual strings. It is also crucial for the operation of databases, not only in sorting records but also in selecting sets of records with fields within given bounds.

It is quite tricky to get collation to work correctly for many languages, and even more difficult to do it with the

speed demanded by customers. Luckily, the ICU library provides a high-performance, full-functioned implementation of international collation, one that is used in IBM products and can be freely used in any other product. This presentation will review the capabilities of ICU collation and illustrate what can be done with it.

Presenter:

John O'Conner
Technical Writer
Sun Microsystems

[Top]

Track 2 - Character Conversions from Browser to Database: or There and Back Again
This presentation considers the various points of character set conversions that may occur in a typical web based application. The session describes the path of character data as it travels from browser form, through 2nd tier business logic, and to its brief resting place within a database. Then it follows the data on a return trip to the browser. All along, common pitfalls and conversion errors will be described.

Presenter:

Han-Yi Shaw
Lead Program Manager Macintosh Business Unit Microsoft Corporation

Top

Track 3 - Unicode in Office for Macintosh: Past, Present, and Beyond
Microsoft Office 2004 for Macintosh represents the first "Unicode-Throughout" version of Office for Mac with significantly improved character fidelity and layout compatibility. This presentation will provide a history on the evolving customer demands for Unicode support in Office for Macintosh and how the teams at the Macintosh Business Unit have endeavored to deliver on the promise of a "Unicode-Throughout" suite of application with the release of Office 2004 for Macintosh. Additionally, the presenter would also like to exchange ideas with leaders from the community, and collect feedback on our Unicode efforts for the next version of Office for Mac.

Presenter:

Adil Allawi
Technical Director Diwan Software Limited

Top

Track 4 - Unicode in Education
One of the biggest obstacles to a developing country is the printing and distribution of educational textbooks in its native language. There are many advantages to using computers in place of paper book. Yet converting complex language textbooks to electronic form can be a non-trivial task. My company developed Islamic studies e-books for the Saudi ministry of education and educational electronic text books for several Arab countries. This paper gives a case-study of problems encountered and the solutions that were applied. While the examples given are for Arabic, the issues are relevant to many developing countries.

| 14:30-15:20 | SESSION 4 |

Presenter:

Marin Millar
Globalization Manager Microsoft Corporation

Track 1 - Building Multilingual Websites with ASP.NET 2.0
This talk will cover the new designer features and resource architecture for localization in ASP .NET 2.0. It will show how to generate resources using the designer. It will also explain new methods for detecting the browser language, the declarative localization model, resource extensibility, and multilingual website deployment.

Presenter:

Tex Texin
*Internationalization
Architect
Yahoo!*

Track 2 - 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:

Andy Abbar
*Director
International
Strategic
Initiatives
Microsoft Europe*

Track 3 - Local Language Program
The importance of language drives Microsoft's global product strategy. The Local Language Program is a global initiative to provide localized desktop software and tools to customers through collaboration with local governments and language experts to build a local IT economy. The Local Language Program has three pillars (language/culture, technology and community) and serves to bridge the language and digital divides between emerging and developed markets, preserve the language and culture. I will share the approach, success stories and our plans for Windows Vista and Office2007.

Presenter:

Waris Abdukerim Janbaz
*Researcher &
Ph.D. Candidate
University of Paris
VIII*

Track 4 - Web Development Considerations for Unicode-based Text Processing in Uyghur Language
This presentation focuses on text processing and accessibility of an agglutinative Turkic language -- Uyghur. It describes the basic concept of Uyghur Unicode font developing, character displaying and character inputting methods within Uyghur-support-less environment. This article also highlights the problems caused by the absence of two Uyghur characters in the Unicode Standard's table, and more importantly, the article proposes solutions to overcome such incompleteness.

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

16:00-16:50	SESSION 5

Presenter:

Evan Gerber
*User Experience
Consultant
Molecular*

Track 1 - Ich bin eine Website – The Impact of Language and Culture on Internationalization & Localization
In this session, Evan Gerber draws off of practical experience on multiple internationalization projects to clearly illustrate the process and pitfalls inherent in developing for multiple cultures and languages. Attendees will garner practical insights on the challenges, strategies, and lessons learned which are critical to the success of any internationalization project. The discussion moves through all phases of software development, from team building to ongoing maintenance. A comprehensive overview of the impact of language and culture on user experience and technical architecture ensures that

participants are aware of the big picture impact.
Exploring specific tools in detail provides tangible
knowledge which can be applied immediately.

Presenter:

Richard Gillam
*Software Engineer
IBM Corp.*

Top

Track 2 - A Forest of Tries
Most text-processing applications have to look up
arbitrary-length
character strings in lists, either to test for membership in
a group or to
look up additional information about a particular string.
The trie is a
classic technique for representing collections of strings,
offering a good
combination of memory efficiency and lookup
performance. This talk will
examine the basic concept of the trie and look at various
implementations--
from simple to complicated-- of tries that optimize for
different types of
list and application.

Presenter:

Russ Rolfe
*Lead Program
Manager
Microsoft*

Top

**Track 3 - Windows Vista, An Ever Expanding View of
Internationalization -- Updated**
Windows Vista expands upon the foundation created in
Windows 2000 and Windows XP to support users'
Internationalization needs. In this presentation, we will
introduce the newly supported locales, user input options
and the extended font coverage. Then we will show how
the OS's localized User Interface will be wholly supported
with the Multilanguage User Interface (MUI) technology
introduced in Windows 2000 and improved in Windows
XP. Plus, we will discuss how Language Interface Packs
(LIPs) will be used to broaden the localized language
coverage. Finally the concept of custom cultures will be
presented.

Presenter:

Toshiya Suzuki
*Research Assistant
Hiroshima
University*

Top

**Track 4 - Investigation of Encodings in Legacy Khmer
TrueType Fonts and Possibility of Auto Encoding
Detection**
In spite of the standardization of character encoding
implementation of OpenType based Unicode rendering
system, still legacy character encodings are widely used
for South or South-East Asian scripts to work with
systems lacking intelligent renderer. Although some
scripts have de-jure or de-facto standards, other scripts
have no standards at all. If the document uses such un-
standardized fonts, the document lacks the correct
delcaration of encoding. Taking Khmer script as a typical
example, we investigate the encodings in the freely
distributed Khmer TrueType fonts and propose algorithm
to identify which encoding is used in the fonts.

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

Presenter:

Christian Donner

Track 1 - A Generic Content Localization Taxonomy
If you are planning to go international with your web
presence, or if you are planning to consolidate a number

*Director,
Distributed
Systems
Aviva Life
Insurance
Company*

Top

of local websites into a single international web presence, the translation of your content is a secondary concern. Before you can start translating, you must define how content is structured and where it will live, so that it can be targeted at the right audiences in the right markets. In this session, Christian Donner presents a generic approach for establishing a content localization taxonomy that can be customized for your individual circumstances.

Presenter:

Michael Kaplan
*Technical Lead
Microsoft
Corporation*

Top

Track 2 - Sorting It All Out: 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:

Ienup Sung
*Staff Engineer
Sun Microsystems,
Inc.*

Top

Track 3 - Unicode Transition Planning for Solaris
Ubiquitous Unicode / ISO/IEC 10646 is one of the most successful character set standards that is widely accepted and implemented by numerous modern information technology standards and products. The Globalization group at Sun believes that adopting the Unicode as the default character set at Solaris will benefit our customers and also Sun. This technical presentation and demonstration outlines the latest Unicode Transition Planning for Solaris at Sun and asks for your comment.

Presenter:

Thomas Milo
*President
DecoType*

Track 4 - Tasmeem - Ultimate Arabization
DecoType and WinSoft have joined forces to build the first ever complete implementation of ACE technology, dovetailing it into the Adobe Creative Suite. The result, dubbed Tasmeem (in Arabic meaning both Design and "Tenacity in Pursuing a Project"), realizes everything - and more - that was ever rumoured about DecoType. ACE technology was created from scratch to meet both Arabic and Unicode requirements without compromise. The presentation will consist of a brief summary of the concepts and a demo of the most striking features of Tasmeem and their relevance for Unicode Arabization.

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

Friday, November 17, 2006

09:00-09:45	**KEYNOTE - Where Would We Be Without UNICODE?**

Moderator:

Dr. Richard Soley

In this Keynote Panel, leaders of standards

OMG

Panelists:

Mark Davis
UNICODE
Richard Ishida
W3C
P. J. Plauger
Dinkumware, Ltd.
Erkki Kolehmainen
CSC-Finnish IT
Center for Science

organizations will explain how they rely on the Unicode standard to forge ahead with their own standards, and how the future of Unicode, in internationalization, localization and globalization, will support their own standards programs.

Other industry experts will detail how Unicode is key to enabling new capabilities and relate the business reality of supporting additional languages. Finally, the panel will discuss trends in upcoming standards and new technology that affect internationalization.

09:45-10:00	Morning Refreshments

10:00-10:50	SESSION 7

Presenter:

Mark Davis
President,
UNICODE Consortium

Top

Track 1 - What's New in Software Globalization Standards
This presentation provides an update on the latest developments in software globalization from the Unicode Consortium, summarizing the most important changes in the Unicode character encoding standard, related globalization standards and specifications, the locale data repository (CLDR), etc. It also describes important related developments from the IETF, ICANN, the W3C, and others.

Presenter:

Richard Gillam
Software Engineer
IBM Corp.

Top

Track 2 - A Brief History of Character Encoding
Unicode didn't just spring fully-grown from the forehead of Zeus; it has many historical antecedents and its design is firmly grounded in existing practice. In fact, to understand Unicode fully, it's necessary to understand where it came from. This talk will trace the history of character encoding in information technology from its roots in the Morse telegraph code through to Unicode. It will cover the Morse and Baudot telegraphy codes, IBM's BCD punched-card code, ASCII and EBCDIC, the ISO 2022 encoding scheme, the ISO 8859 family of encodings, and various other encodings, ending with the early history of Unicode and ISO 10646.

[10:00-11:50]

Moderator:
Steven R. Loomis
Software Engineer
Globalization Center
of Competency
IBM Corp.

Panelists:
Mark Davis
Google
Mark Garrett
ModernGigabyte LLC

Track 3 - Panel: Deploying the Common Locale Data Repository
The Common Locale Data Repository is a project for the exchange of
culturally sensitive (locale) information used in application and
system development, and to gather, store, and make available such
data. By pooling resources, the time and expense of collecting good
data can be minimized. As well, minority languages and small
countries have a focal point for submitting data. This session will

Tex Texin
Yahoo!
Ram Viswanadha
IBM Corp.
Daniel Yacob
The Ge'ez Frontier Foundation

go into detail regarding exactly what types of locale information are
available in CLDR, how the data is to be used, and how the CLDR
vetting process works to ensure the quality of data. Panelists will
then discuss how they are making use of CLDR data, the latest project
status, and 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.

11:00-11:50	SESSION 8

Presenter:

Charles Hornig
Globalization Architect
IBM Corp.

Top

Track 1 - Advanced Java Globalization

This presentation describes how to improve the level of global support of Java™ code from "works most of the time" to "works all of the time". The focus is on some specific areas we have identified in reviews of existing code. For each topic, examples of how to code to avoid problems and how to test for them will be provided. The presentation is intended for anyone responsible for producing Java applications to be used in a multinational environment. Attendees should be familiar with the Java programming language, basic globalization concepts, and basic Unicode concepts.

Presenter:

Michael Alford
Senior Architect
Travelport, Inc.

Track 2 - Internationalization of the Travelport B2C Platform

Travelport's U.S.-only travel commerce platform that runs Orbitz.com and CheapTickets.com will be internationalized in 2006 to support the hosting of many brands in different locales and currencies simultaneously on the same server instances. We investigated dates, times, time zones, currency, distance, string comparisons, character sets, geography, white-labeling, and e-commerce concerns for this Java-based platform. This case study of the first phase of the project (the deployment of Ebookers.co.uk) will discuss the key technology choices made during the project, evaluations of Java and open-source internationalization components, and enforcement of best practices within application code.

Track 3 - Panel: Deploying the Common Locale Data Repository
(Cont'd)

12:00-13:00	LUNCH

13:00-13:50	SESSION 9

Dale Schultz
Globalization Test Architect
IBM Corp.

Track 1 - The Why, Who, When and How of Efficient Globalization Testing

A critical part of software development is the testing of the product. Testing the globalization aspects of software is challenging because test staff may not be familiar with the details of the many functional requirements of globalized software. This presentation will describe the new approach being used by IBM to enable test teams to determine what needs to be tested for adequate coverage and how the testing can be performed in an efficient manner. It will describe the best practices for who should do the testing and when. It will also identify some common techniques that should be avoided.

Mark Davis
Google

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.

Fred Susi
Senior Product Manager
John Fay
Senior Product Specialist
Zebra Technologies Group

Track 3 - Embedding Unicode Support Directly into a Printer: A Case Study

The Unicode Standard unified how to encode characters in computer equipment. However, it is left to a rendering device to determine the size, orientation, and placement of glyphs. The rendering device may even be required to choose between a series of glyphs. Furthermore, with many pre-Unicode applications/systems still in place, an output device (printer) must be able to interpret and switch between the various Unicode and Legacy encoding schemes to be commercially successful. This presentation will discuss the issues and solutions uncovered while developing one of the first Unicode-enabled printers that can transform encoded data streams into human readable output.

Moderator:
Deborah Anderson
UC Berkeley

Panelists:
Daniel Yacob
Director
Ge'ez Frontier Foundation
David Germano
Associate Professor
University of Virginia
Leonard Muellner
Professor

Track 4 - Panel: Unicode and the "Lesser Known" Scripts: Historic and Minority Scripts in Unicode, New Developments and Requirements

As the universal character set, Unicode aims to cover the historic scripts of the world as well as modern scripts. The historic scripts are of interest to students and scholars, and also play a critical role in preserving our world cultural heritage. Many are used in modern liturgical contexts. This panel will highlight the new directions Unicode is providing for currently encoded scripts, addressing such questions as: how has Unicode influenced teaching and publishing? Has Unicode made the material more accessible to a wider audience?

Brandeis University
Neel Smith
Associate Professor College of the Holy Cross

What outstanding problems and issues remain? For the over 40 unencoded minority scripts, two questions will be raised: why is encoding these scripts important and how the government can assist.

14:00-14:50	SESSION 10

Presenter:

Charles Hornig
Globalization Architect IBM Corp.

Top

Track 1 - Globalization Testing using Eclipse™
This presentation describes how to use the Eclipse Test and Performance Tools Platform (TPTP) to build globalization tests. TPTP is an open-source platform for testing tools. Among many other capabilities, it provides frameworks for both static and dynamic code analysis. IBM has used these frameworks to build a suite of tests to detect common globalization coding errors. This presentation will describe how this was done and how you can build tests tailored to your own needs. The presentation is intended for those developing globalization testing plans and tools. Attendees should be familiar with basic globalization and Unicode concepts.

Presenter:

Jennifer DeCamp
Foreign Language Technology Program Manager MITRE

Top

Track 2 - Implementing Multilanguage Databases
While Unicode has significantly simplified the problems of storing and searching multiple scripts, the use of Unicode is often not sufficient for many applications, particularly applications involving extensive legacy data. There are applications in which it is necessary to know the language as well as the script, but where language identification software is not effective. There also are issues with languages for which Unicode encodings are not yet available, display systems for languages with complex scripts, Unicode-encoded fonts, and input methods so that users can see and input data. Jennifer DeCamp will discuss the issues and present solutions that are being used by the Board of Geographic Names in their move to providing data in multiple languages and scripts.

Presenter:

John Harvey
Engineer Apple Computer

Top

Track 3 - International Features of Mac OS X
Mac OS X is a modern, robust, Unix-based operating system. This session covers the international capabilities of Mac OS X primarily from an end user perspective, with a particular emphasis on new features in Mac OS X 10.4 Tiger, the latest version. Topics covered include supported languages, input methods and keyboard layouts, locales, font technologies, and user customization. Topics of interest to software developers and language experts will also be considered.

Track 4 - Panel: Unicode and Minority/Historic Scripts; New Directions (Cont'd)

14:50 – 15:10	Afternoon Refreshments

15:10 - 16:00	SESSION 11

Presenter:

Addison Phillips
Internationalization Architect
Yahoo!

Top

Track 1 - The Theory and Practice of Pseudo-Translation
Pseudo-translation is the use of a program to alter software resources "as if" they had been translated. Usually this takes the form of taking unaccented English letters and replacing them with non-ASCII characters. Pseudo-translation can be used to test the "localizability" of a piece of software or the adaptability of the display to non-English characters, fonts, and so forth. This session will present the various uses that an organization might consider for pseudo-translation, along with the in-depth examination of a pseudo-translation system developed by the author. Some of the use cases covered will including visual and functional testing, generating "global" test cases from English-centric cases, and character encoding test generation.

Presenter:

Nelson Ng
Chief Globalization Architect
eBay Inc.

Top

Track 2 - A Case Study of eBay UTF-8 DB Migration
The eBay Marketplace is a global trading platform – 193 million people in more than 150 countries are registered eBay members. The eBay Marketplace has a site presence in 33 markets around the world, 24 of which have country specific marketplaces supported by a single code base that supports eBay members worldwide. This presentation will cover the migration of the eBay Marketplace database from ISO Latin 1 to UTF-8. Nelson Ng, Chief Globalization Architect at eBay, will cover the inline approach used by eBay to address the challenges of migrating over 100 databases with 18+ Billion textual data records while keeping the entire site online. In addition to the preparation and execution of this migration approach, Nelson will cover alternatives considered and lessons learned.

Presenters:

Naoto Sato
Java Internationalization Engineer
Sun Microsystems
Craig R. Cummings
Principal Software Engineer
Oracle Corporation

Top

Track 3 - New Internationalization Features of the Java Platform - Java SE 6
See what internationalization features are planned for the next version of the Java Platform -- Java SE 6 (codename Mustang). Highlighted in this session will be features such as Locale Sensitive Services SPI, Normalizer API, ResourceBundle enhancements, and new Japanese calendar support. It also gives you a sneak peek of the latest planned features beyond.

Presenter:

Track 4 - Chinese Dilemma: How Many Ideographs are Needed?

Kuang-Hui Chiu
*Associate Professor
National Taipei
University*
Chi-Ching Hsu
*Standard and
Business Planner
IBM Taiwan Corp.*

Top

A Chinese proverb says that, "great quantity is equal to beauty". This presentation will apply this adage and describe what scenario may occur if the proverb is applied to the information technology industry and the question of how many ideographs are needed. It will explain the dilemma of whether to accept the niche market requirement to collect and display all ideographs or follow the more common users' requirement to provide a basic set of ideographs only.

16:10 - 17:00   **SESSION 12**

Presenter:

Pierre Cadieux
*President
i18N Inc*

Top

Track 1 - The Localization Process and Applicable Standards
A high-level architectural model of the localization process is constructed, describing the steps required for scalable and sustainable multilingual deployment. The model is built graphically, item by item, starting from source content creation through change detection, text extraction, leveraging, workload, translation, review, etc. to produce localized content in several languages. The main characteristics of each step, and applicable software components, are discussed. The model then serves as backdrop to discuss which localization standards exist (or are under development) to support the localization process. You will learn where and how standards such as JCR, XLIFF, TMX, TWS, GMX/V and others, interact.

Presenter:

Charles Hornig
*Globalization Architect
IBM Corp.*

Top

Track 2 - Overhauling a C++ Application to Use Unicode
This case study describes the process of incorporating Unicode functionality into an existing multi-platform C++ application, from design through application delivery. The focus is on the technical challenges and how they were met (or not). The topics discussed include: compatibility with existing data and interfaces; divergent Unicode models on different operating systems; dealing with relational databases; dealing with third-party libraries; the operating system interface; choosing what to test. The presentation is intended for those considering globalization improvements to existing applications. Adding Unicode functionality can add years to the life of an application at much less cost than a complete replacement.

Presenter:

Michael B. Toth
*Program Manager
R.B.Toth Associates*

Doug Emery
*Data Manager
Emery IT*

Top

Track 3 - From Parchment to the Web: Integrating and Encoding the Archimedes Palimpsest Transcriptions
An international team has been digitally imaging, transcribing and studying the 1,000-year-old Archimedes Palimpsest (http://www.archimedespalimpsest.org/). Integrating transcriptions of Archimedes' mathematical texts with multispectral digital images and hosting them on the Web for global users poses a complex set of

information sharing challenges. Unicode is proving to be a key component of this effort as a standard encoding scheme that conveys the sense of the scholars work clearly and efficiently to readers using a range of software to access the data. It is also an essential element in the integration of the transcribed information with the archive of complex digital images.

Presenter:

Su Liu
Cheng Xu
Advisory Software
Engineers
IBM Corp.

Top

Track 4 - Build GB18030 from Unicode Enablement Features
GB18030 enablement is important because it is a mandatory national
standard in China. To simplify development of compliant systems, GB18030
support can be built on existing Unicode features and functions. This presentation will cover the following aspects of how to manipulate GB18030 in Unicode enabled systems: 1) Unicode features for Chinese support, 2) Basic
requirements for GB18030 implementation, 3) Enablement features of GB18030
and Unicode; a. locale definition, b. input method, c. printing, d. font;
4) Solution to GB18030 enablement on Unicode features, 5) Character
conversion between GB18030 and Unicode, 6) Handling 4 byte GB18030
characters.

Program is subject to change.

- To Register for IUC30:
  http://www.unicodeconference.org/registration.htm
  Or, contact Kevin Loughry at loughry@omg.org

- Exhibitor Information: http://www.unicodeconference.org/be-exhibitor.htm
  Or, contact Ken Berk at ken.berk@omg.org,

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

- Early-bird registration discount deadline: October 2, 2006

- IUC30 will be held at the Hilton Washington Dulles Airport, Herndon, VA, USA November 15-17, 2006