



Internationalizing the Virtual World

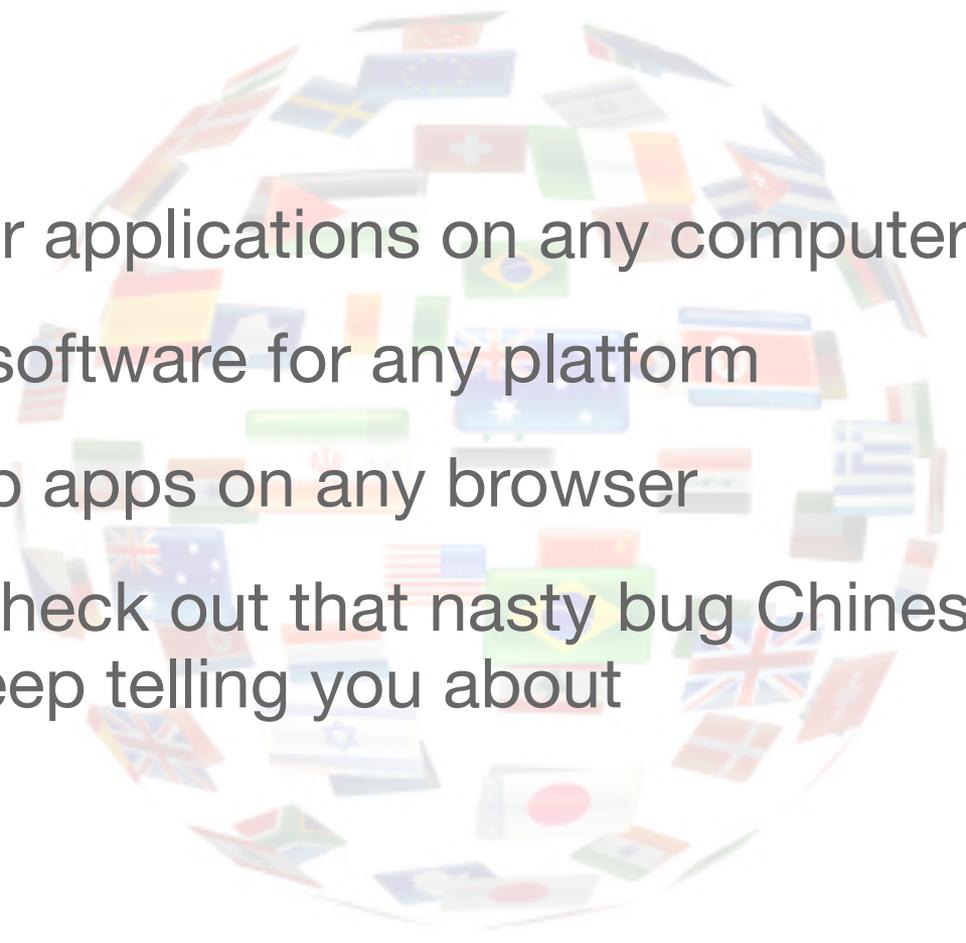
Ben Gertzfield
Senior Member of Technical Staff
IMUG, 2008-07-17

“

I know my i18n, l10n, and g11n...

What's this v12n thing?

Virtualize World Peace

- 
- Use your applications on any computer anywhere
 - Design software for any platform
 - Test web apps on any browser
 - Finally check out that nasty bug Chinese Windows users keep telling you about

Why Virtualize on Mac?

- Develop and test your software:
 - On every operating system
 - In every language
 - All from the comfort of your Mac



Break down the language barrier

- Distribute software and its OS in a virtual appliance
 - No more software that only runs on Japanese Windows
 - Run any software on Windows, Mac, and Linux
 - No installer or setup needed
 - No dependencies on system libraries or device drivers

VMware Fusion: Lots of Firsts

- First Mac product developed at VMware
- First consumer product
- First Unicode product
- First release in English, Japanese, French, and German



Virtualization as the Univeral Translator



- Integrate Unicode and legacy OSes
- Copy-and-paste styled, international text
- Shared Folders link Mac and Windows filesystems
- Drag-and-drop filenames with different Unicode normal forms

Unicodefication

- VMware: Unified source code base in C, C++, IA-32 assembly, Objective-C, Objective-C++, Java, C#, Python, Perl...
 - Before Fusion: no internationalization framework
 - All strings blindly passed as byte arrays (`char *`)
- Unicode and ICU to the rescue!

ยูนีโคด

יוניקוד

유니코드

يونيكود

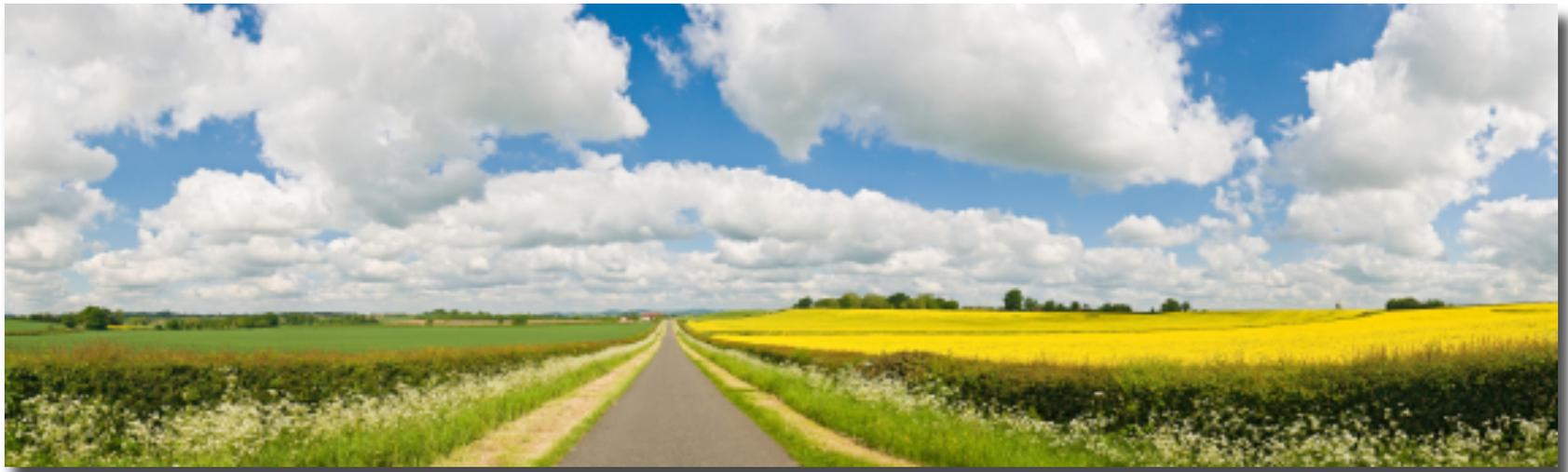
Unicode

One Million Calls to strtolr

- Initial i18n redesign used UTF-16, required ICU everywhere
 - ICU's C++ runtime dependency makes Linux distribution extremely difficult
- Refactored out pure C encoding conversion routines from ICU
- Implemented cross-platform UTF-8 safe string library

The Long Road to Unicode

- Mac OS X is (mostly) Unicode
- No encoding had been ever defined for VMware file formats and protocols
- No cross-platform standard for locale information



Windows: No UTF-8 For You

- Cross-platform code base required compromise
 - Linux filesystems and APIs have no defined encoding
 - Windows provides UTF-16 versions of most APIs
 - Mac OS X string APIs support UTF-8 and UTF-16
- POSIX and Win32 API wrappers to convert UTF-8 to native platform encoding

Lots of Messages in Lots of Bottles

- Background processes pass localized and unlocalized messages via RPC
- Extracting messages from Mac, Linux, Win32 binaries
- NIBs and .strings files for Cocoa-only apps



One Mac, Many Languages

- ADB keycodes undocumented, some values swapped on European keyboards
- No AltGr, no zenkaku, no hankaku
- Key remapping helps, but Mac-like input methods are better

“

**Time is an illusion.
Virtual time doubly so.**

(With apologies to Douglas Adams)

Making Time Work



- Easy Install: Hands-free installation of virtual machines (including time zone configuration)
- Windows time zone API uses hard-coded set of time zones
- Mapping POSIX time zones to Windows: GMT offset plus judicious guessing for DST

- VMware Workstation 6.5 first fully internationalized release
- VMware Fusion 2.0 now opens legacy non-UTF-8 VMs created on Windows or Linux
- VMs with non-US ASCII names and data fields portable between Mac, Windows, and Linux for the first time

“

Time for Q&A!

请你们提问。

ご質問などありますか？

Haben Sie Fragen?