CFUnited 2008
Washington D.C.

Oğuz Demirkapı

CTO
NicheClick Media
http://NicheClick.com

MULTI LINGUAL (I18N) COLDFUSION APPLICATIONS

About Me

- CTO at NicheClick Media

- Coding since '85, CGI Programming since '94

- ColdFusion Developer since '97 (Flex since '07)

- Founder and Manager (prev) of CFUG for Turkey

- ~~Team Fusebox Member~~

- Interested in
  ColdFusion, Flex, AIR, Ajax, Frameworks, i18N, L10N, G11N

- Have big interest in Epistemology

- Living in Potomac, MD, moving to Irvine, CA

- Personal blog: http://blog.demirkapi.net

Agenda

- What you may expect?
- i18N Theory & Basics
- ColdFusion & i18N
- i18N Tips (Code & Server & Database)
- Java Resource Bundles
- Application Modeling
- Resources
- Questions & Answers
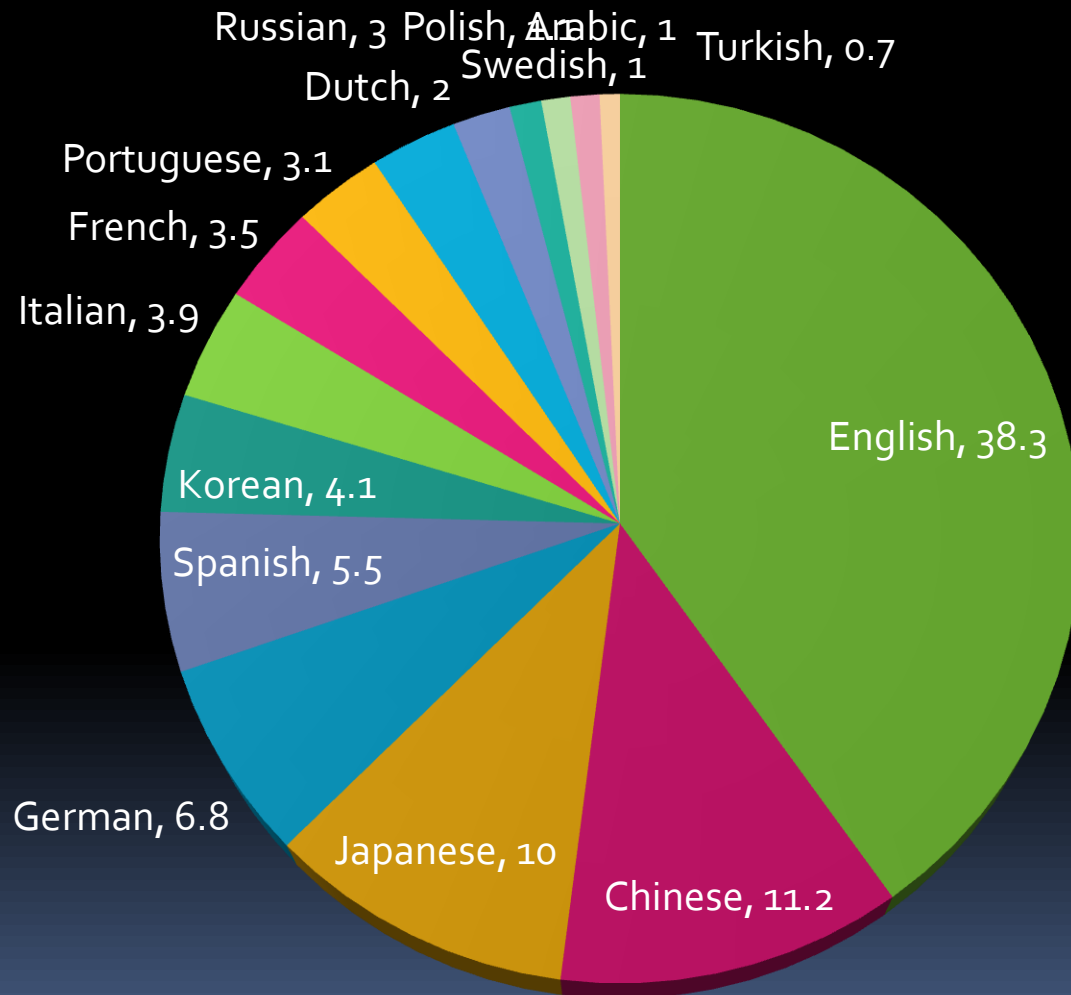
What You May Expect?

Open Your Eyes, if;

- You are a i18N newbie
- You heard that you should create i18N applications
- You need to get idea about i18N usage
- You want to get some tips

Attend another session, if;

- You are already using i18N in ColdFusion in production
- You think I know everything about i18N implementation ☺

Do we need multilingualism?

- Yes!

Pie chart of language distribution:
- English, 38.3
- Chinese, 11.2
- Japanese, 10
- German, 6.8
- Spanish, 5.5
- Korean, 4.1
- Italian, 3.9
- French, 3.5
- Portuguese, 3.1
- Russian, 3
- Dutch, 2
- Polish, 1.2
- Arabic, 1
- Swedish, 1
- Turkish, 0.7

Problems

- Charset Conversions
- Formatting & Parsing
  - Date & time
  - Messages
  - Numbers & currencies
- Translated Names
  - Languages, Regions (Countries), Scripts, Time zones, Currencies
- Calendar, Time Zone, Date/Time conversions

- Collation
  - Searching, Sorting, Matching
- Segmentation
  - word, line, …
- Transforms
  - Normalization
  - Casing
  - Transliterations
- Unicode Regular Expressions
- Complex-Text Display / Input
- …

What is …?

- ## What is i18N (Internationalization)?

  - ### Application functions in at least two locales

- ## What is L10N (Localization)?

  - ### Process of applying a locale or language "skin" to an i18N application

- ## What is G11N (Globalization)?

  - ### i18N & L10N

Globalization & Localization

Globalization

- Single character set
- Single executable
- Single install
- Single server serves all clients in all languages

Localization

- Based on globalized software
- Adds specific translations and adaptations for particular languages and markets

Globalized software can be localized without any code changes

Why i18N (Internationalization)?

- The World Wide Web should be truly world-wide!
- Internationalization is important to ensure that users worldwide can equally benefit from Web technology
- Wide diversity world-wide:
  - Scripts (Latin, Cyrillic, Hebrew, Tamil, Katakana,…)
  - Languages (English, German, Turkish, Korean, Japanese,…)
  - Typographic conventions
  - Cultural conventions
  - Political circumstances
- Avoid fragmentation of specifications due to localization
- Make sure internationalization is done at the right place

What is Locale?

- The combination of a language and a country code

  en_US: US English (color, $)

  en_GB: British English (colour, £)

  de_CH: German in Switzerland

  tr_TR: Turkish in Turkey

Selecting Locale

- Manual
  - Default locale can be loaded and other options would be available by selection via buttons/selects etc.
- Locale Detection
  - Capabilities.language property in Flash Player and AIR
  - Parsing 'Accept-Language' header on HTTP request
    - Not usable with URLLoader
  - Parsing browser and OS language settings via JavaScript
  - Location detection depending on IP

Character encoding

Character encoding specifies mappings from a character set to the integer numbers that represent the characters on a computer.

- EUC-JP (Japanese)
- EUC-KR (Korean)
- ISO-8859-1 (Western European and English)
- SHIFT_JIS (Japanese)
- UTF-8 (All Languages)

What is Unicode?

- Unicode (unicode.org) is an character set for all the characters and symbols of the world.
- Unicode provides a unique number for every character.
- Except Klingon! ☺

ما هي الشـفرة الموحدة "يونِكود" ؟

Unicode nedir?

Τι είναι το Unicode

Cos'è Unicode?

유니코드에 대해?

Что такое Unicode?

Unicode (cont.)

- **Why do we need to use Unicode?**
  - Avoids data corruption
  - Single encoding for text in all languages
  - Makes software globalization possible
    - Vastly reduces development cost
    - Vastly reduces maintenance, update and support cost

- Switching to Unicode has no disadvantages for single language users, to the contrary it usually offers advantages even for single language users. And it offers great advantages for multilingual users.

- Encoding: Use Unicode wherever possible for content, databases, etc. Always declare the encoding of content.

Unicode (cont.)

- Non-Globalized Component
  - Does not use Unicode
  - Hard-coded date/time formatting & parsing
  - Hard-coded number & currency formatting & parsing
  - Hard-coded collation (sorting/searching/matching)
  - Other hard-coded operations
  - Hard-coded literals

Unicode & Files

- UTF-8 is the recommended encoding for files.
- Use a Unicode capable editor (IDE)
- Flex Builder
  - Default encoding is UTF-8
- Eclipse
  - Default encoding is Cp1252
  - Change it into UTF-8
    - Window-> Preferences -> General -> Workspace
      - Text file encoding
- Dreamweaver
  - Supports both Unicode and BOM
- Do not use Homesite/CFStudio
  - http://www.adobe.com/go/tn_19059

ï»¿<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

Unicode & Database

- Use a robust database with right settings
  - MS SQL Server
  - MySQL Server
  - PostgreSQL Server
  - Oracle
- Beware of Unicode Support
  - MySQL 4.1 and up

    default-character-set=utf8
  - MS SQL Server
    - SQL Server nvarchar, ntext etc.
  - MS Access 2000 and up

ColdFusion & i18N

- ColdFusion Versions
  - ColdFusion MX
    - New Java platform features
    - Java Libraries (such as IBM ICU4J library)

  - Converting ColdFusion 5 or older projects into MX is a headache with Unicode

ColdFusion & i18N – Versions

- ColdFusion 5
  - No native Unicode support
- ColdFusion MX
  - Lack of MS Access Unicode support
- ColdFusion MX 7
  - Advances Java locale supports
- ColdFusion 8
  - Updated locales

ColdFusion & i18N – Server

- Default character encoding is Unicode
  cf_root/lib/neo-runtime.xml
  ```
  <var name='defaultCharset'>
    <string>UTF-8</string>
  </var>
  ```

- Use ColdFusion MX
  - If possible ColdFusion MX 7.x and up
- Use Updated JVM
- Setting Correct Locale
  - ColdFusion does not start when locale is Turkish (tr_TR)

ColdFusion & i18N – Code

- ## Use Unicode
  - If possible with BOM (byte order mark) support
- ## ColdFusion MX Detect BOM
  - If you have BOM support on your file ColdFusion understand your locale and there is no need to use extra tags such as cfprocessingdirective
  - If there is no BOM, in every single ColdFusion template

    <cfprocessingdirective pageencoding="utf-8“>

    Using in Application.cfm etc. does not help

ColdFusion & i18N – Charset

- ## Beware of ColdFusion Functions & Tags

  CFPROCESSINGDIRECTIVE, CFCONTENT, CFFILE, CFHEADER, CFHTTP, CFHTTPPARAM, CFMAIL, CFMAILPART tags and the
  SETENCODING, GETENCODING, TOBASE64, TOSTRING, URL DECODE, and URLENCODEDFORMAT functions etc.

- ## Use encoding in Tags

```
<cfmail to="#user#" from="#me#" subject="Unicode Test" charset="utf-8">
 <cfmailpart charset="utf-8" type="plain">#mymailasplaintext#
    </cfmailpart>
    <cfmailpart charset="utf-8" type="html">#mymailashtml#
    </cfmailpart>
 </cfmail>
```

ColdFusion & i18N – Web

- HTML/XHTML

  <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

- Multipart POST

  <form action="#FOO#" method="post"

  enctype="multipart/form-data: charset=utf-8">

- Detect Locale

  - cgi.HTTP_ACCEPT_LANGUAGE

  - IP & GeoLocator

ColdFusion & i18N – Database

- **MS SQL Server**
  - Use ColdFusion Admin DSN Settings for Unicode Support and CFQueryParam together

  OR

  - Use 'N's
    http://www.adobe.com/go/tn_18331

- **MySQL**
  - 3, 4, 5 versions supported natively

- **Access**
  - Use Unicode drivers in ColdFusion Admin

Code & Language Isolation

- MVC Benefits
- Java Resource Bundles
  - JRB General usage with CFCs
  - IBM ICU4J Library
- GNU GetText ()
- XML
- Database usage
  - Language flagged content for different languages
  - Prefixed database tables

Java Resource Bundles

- ColdFusion has resource bundle tools
  - Use it!

  Thanks to Paul Hastings

  http://www.sustainablegis.com/blog/cfg11n/

- Go beyond and install IBM ICU4J Library for more i18N support

  http://icu-project.org

  - Provides extended i18N functionality

Java Resource Bundles Tools

- Attesoro

  http://attesoro.org

- Eclipse Plugin
  http://sourceforge.net/projects/eclipse-rbe/

- Web based

  - http://www.jmpj.net/rbMan/

DEMO

Base Resource Bundle Usage

- Based on CFCs
- Contains different kind of methods
  - getResourceBundle
  - loadResourceBundle
  - formatRBString
  - messageFormat
  - Etc.

DEMO

Application Modeling

- Architecture
  - Same content in every locale
    - Every template has identical content in different language
    - Suitable for portal style applications
  - Different content tree in different languages
    - Suitable for content based sites

Resources

- Adobe Developer Center

    http://www.adobe.com/devnet/coldfusion/localization_globalization.html

- Globalizing ColdFusionMX

    Paul Hastings http://www.sustainablegis.com/blog/cfg11n/

- Unicode http://www.unicode.org

- IBM software globalization

    http://ibm.com/software/globalization

- My blog

    http://blog.demirkapi.net

Questions & Answers

?

Oğuz Demirkapı

http://blog.demirkapi.net

cfunited@demirkapi.net