Corporate Globalization Team

The Globalization Architecture Imperatives

Corporate Globalization Team

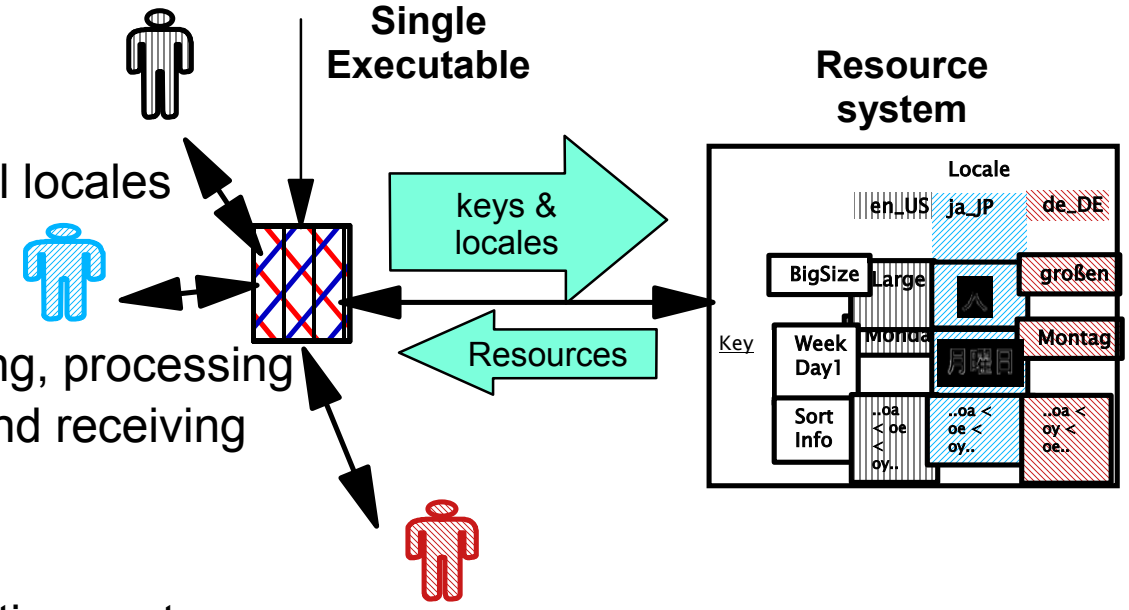
GAI and the IBM Software Strategy for e-business

- Globalization needs to be designed into the product from the very beginning
  - Clients can be in any language, located anywhere worldwide
  - The firewall need to be able to handle users correctly (name in different languages)
- The web application server needs to:
  - Provide information in the correct language, format based on client requirements
  - Be connected to external services (various languages and code pages)
- The Infrastructure services must manage entities in different languages

Corporate Globalization Team

Single source, single executable ...

- ... allows data & code separation ...
  - ... and thus a single executable for all locales
  - Devoid of any cultural information
  - Only useful if can perform:
    - Data entry, display, printing, editing, processing
    - Storage, retrieval, transmission and receiving



- Benefits
  - Reduce overall development and testing costs
  - Quicker time to market
  - Reduce maintenance cost
    - Mix and match components, addition of new languages
    - A single server can support clients in any language

Corporate Globalization Team

Multilingual data

- Enable single software products or e-business applications to be targeted across multiple platforms, languages and countries without re-engineering
- Allow data round-tripping without corruption
- Programs part of an e-business system should be Unicode based
- Integrating legacy applications part of e-business systems require special code, called Connectors
- Unicode is the universal character encoding scheme for written characters and text
  - Can be used as the lingua franca on the server and database even when the client uses legacy encoding systems

Corporate Globalization Team

International Components for Unicode (ICU)

- Open source and cross platform library to support Unicode
  - Same set of functions available in Java (most incorporated by Java) and C/C++
- Support for working with Unicode strings
  - Collation, iteration, character classification
- Efficient conversion to/from other encodings
- Follows the Unicode standard as it evolves
- Java, XML and HTML are Unicode based

- All IBM product groups are committed to making ICU services available on their platforms

- Web Site: http://oss.software.ibm.com/icu/

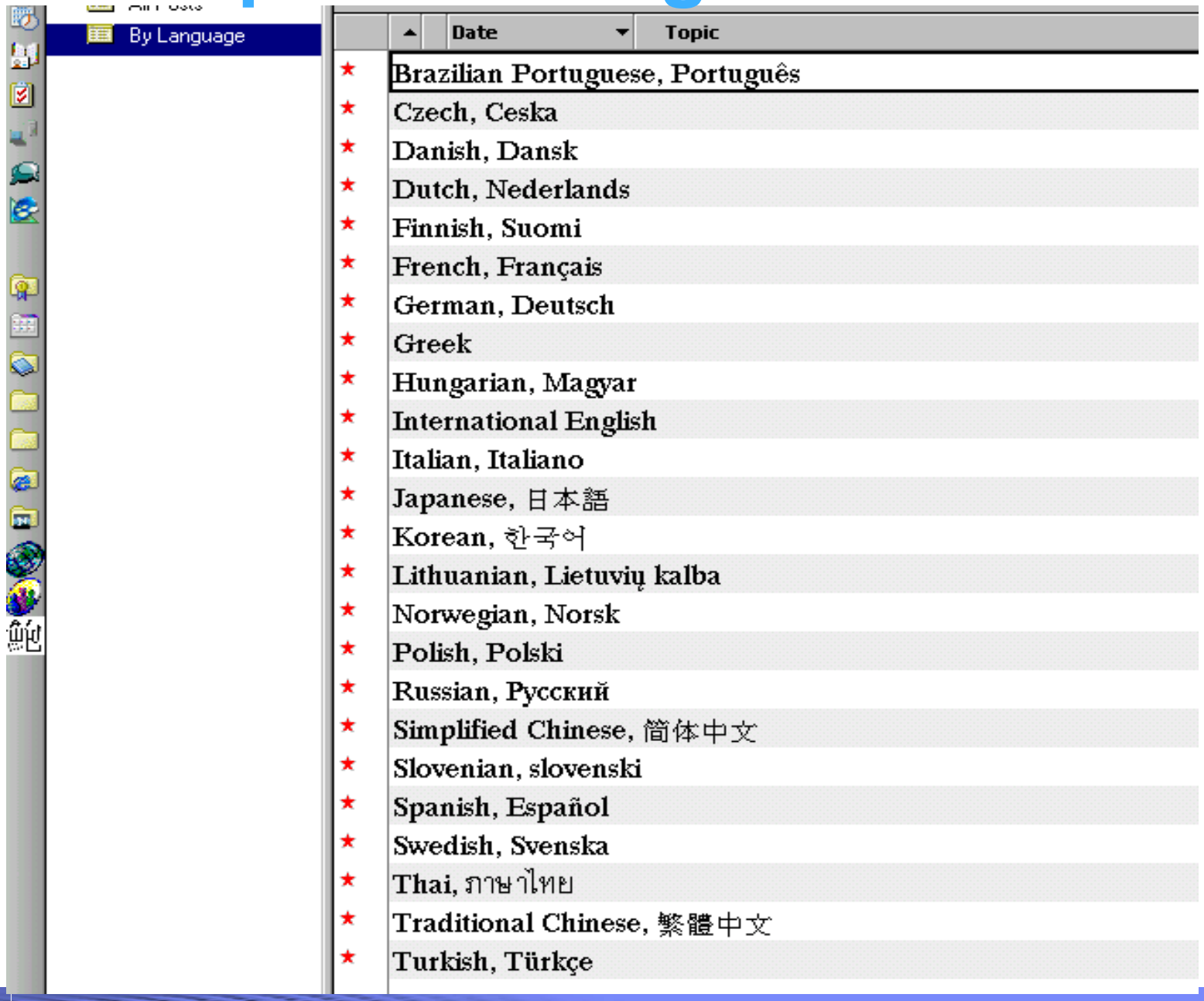
Corporate Globalization Team

The Domino/Notes experience

- Notes uses LMBCS (Lotus Multibyte Character Set)
  - Defined before Unicode
  - Data in standard code page with a LMBCS group identifier
  - Multilingual data support
- Extensive customer base already using LMBCS
- Use one LMBCS group for Unicode
- New languages (e.g. Hindi, Tamil) implemented as Unicode
- Internally developed library for data handling
- Migrating to ICU in Domino 6
- Use Unicode API for display/print
- Same character in different code pages treated as identical
- Search by Unicode or LMBCS

Corporate Globalization Team

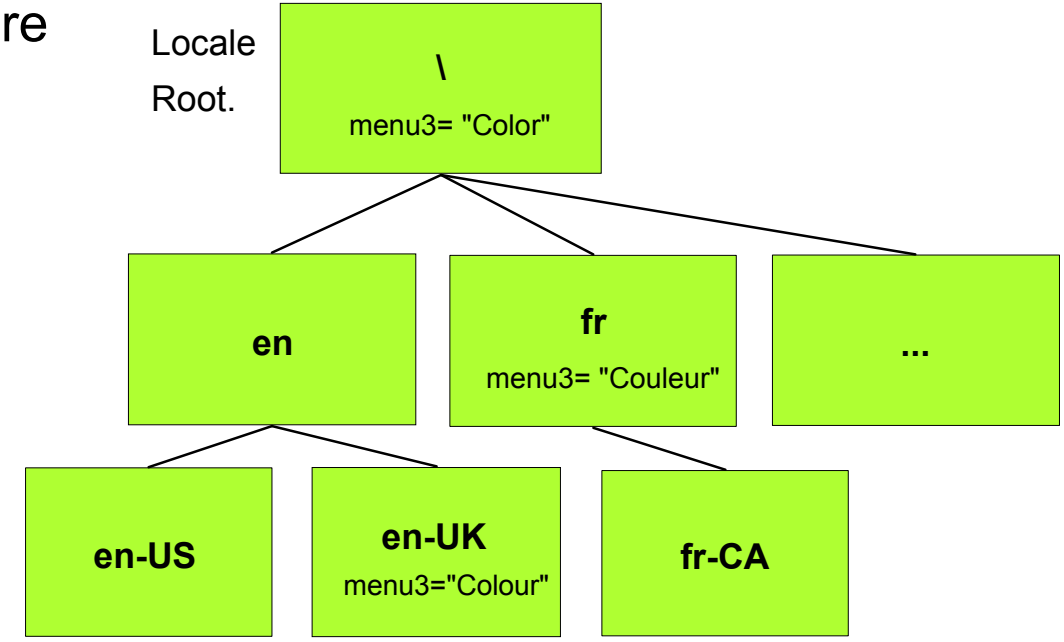
Sample multilingual Notes



Corporate Globalization Team

The "locale" model

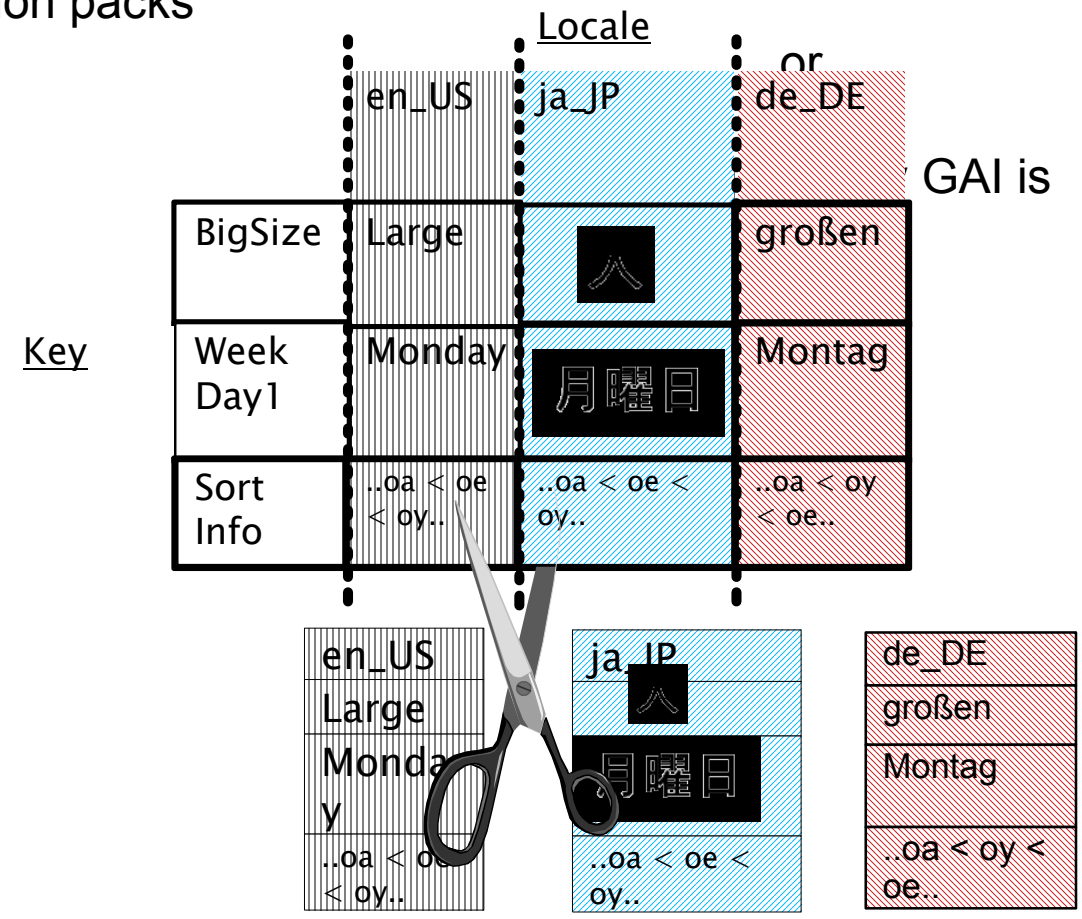
- A locale is a specification of "a language and country" or "a language, country and variant"
  - English-Canada, French-Belgium
  - "en_US" for English (U.S.)
    - ISO 6392 - for language. ISO 3166 for country codes
- A simple way for different parts of an e-business system to summarize and communicate the varieties of culturally-expected behavior
- POSIX and Web (RFC 1766) models are both reasonable starting points
  - Both systems string-based
- Human readable and expandable



Corporate Globalization Team

Localization packs

- Required to support a Single Source, Single Executable model
- Data specific to locales need to be separated from the executable
  - Contained in one or several localization packs
  - They can be Installable, Pluggable On-demand
  - The minimum support Pluggable



Corporate Globalization Team

Input/Output of multilingual data

- World languages include many different kinds of scripts or writing systems
  - Different requirements for input and output
  - Complex Input
    - Asian languages require IME's
  - Complex Output
    - Arabic and Hebrew are displayed Right to Left
      - Interspersed with other script that display Left to Right
  - Different challenges with printing
- Established standards to handle the input, output, storage and display of scripts
- IBM provides Unicode fonts with most of its products
  - Licenced from Monotype - WorldType

Corporate Globalization Team

Additional requirements

- Requirements for common e-business needs, not required for all elements of e-business applications
  - Cross platform, cross application & cross industry building blocks
  - "Intelligent infrastructure"
- Linguistic Services
  - Work with human semantics of the language data in an e-business system
  - Grammar checker, dictionaries, Text to Speech, Speech to Text, ...
- Legacy data and applications
  - Integrate the vast amount of information store in legacy systems
  - How can it be preserved while interacting within a Unicode environment?
- Business objects
  - Encapsulate common business operations to be shared across applications and work across locales

Corporate Globalization Team

Linguistic services

- For many applications, globalization means character handling
  - No imapct on the meaning of words and sentences
- Linguistic Services require extensive understanding of supported languages
  - Low level linguistic tools
    - Spell Checker, Hyphenation, Thesaurus, Morphological analysis, Grammar Chekcer, Disambiguation, Segmentation, Dictionary
  - Data mining
    - Text categorization, text summarization
  - Speech
  - Text Search
  - Machine Translation
- Some services are not available for specific languages
  - Hyphenation has no mean in Chinese, Japanese, Korean and Thai
- IBM has been a leader in linguistic research for years

Corporate Globalization Team

Legacy data and applications

- According to some research, 80% of the business world most valuable data is stored in legacy applications
  - IBM is improving its e-business connectors to be global
    - Pervasive use of Unicode
    - Support of single executable and localization packs
    - Support for and/or exploitation of the text processing functions described in the Unicode Standards
      - Transformation from logical to visual order of bi-directional data
      - Processing of complex script languages
      - Locale sensitive normalization and collation
- Migration strategies for globalization
  - Conversion of existing data to Unicode
  - Recognition of the limitations of some OS/390 data stores (e.g. VSAM KSDS) capabilities to effectively store and retrieve Unicode data and offer guidance on alternatives
  - Unicode-enablement of existing applications, like most COBOL programs, with limited current internationalization capabilities and others, like those written in C/C++, that might be highly internationalized, but are not globalized

Corporate Globalization Team

Global business objects

- Business objects encapsulate commonly used business functions into EJB's
  - Shared and re-used across applications
- They need to be designed to support different languages and provide functions based on local conventions
  - Need to follow a localization pack architecture
- IBM has a very extensive experience helping customers all over the world implement such applications

Corporate Globalization Team

Supporting processes

- Globally aware development tools and processes
  - Creating and maintaining global e-business applications require a set of inter-operable tools and processes
- Certification and review
  - IBM is tracking score cards to certify all platforms and software
  - Compliance to the requirements of Single executable, Multilingual data, Locale model and localization packs
- Localization
  - Process of creating localization packs
    - Application
    - Middleware
    - Operating Systems or other components

Corporate Globalization Team

Спасибо
Russian

Merci
French

Gracias
Spanish

감사합니다
Korean

شكراً
Arabic

תודה רבה
Hebrew

धन्यवाद
Hindi

多謝
Traditional Chinese

Tack så mycket
Swedish

Obrigado
Brazilian Portuguese

go raibh maith agat
Gaelic

Tak
Danish

Grazie
Italian

多谢
Simplified Chinese

Danke
German

ありがとうございました
Japanese

Dank u
Dutch

ขอบคุณ
Thai

Thank You
English

நன்றி
Tamil

Dekujeme Vam
Czech