

White Paper



Designing Global Applications: Requirements and Challenges

Sourav Mazumder

Abstract

This paper explores various business drivers for globalization and examines the nature of globalization requirements for software products and applications. It discusses the typical globalization challenges that must be addressed before building global applications.

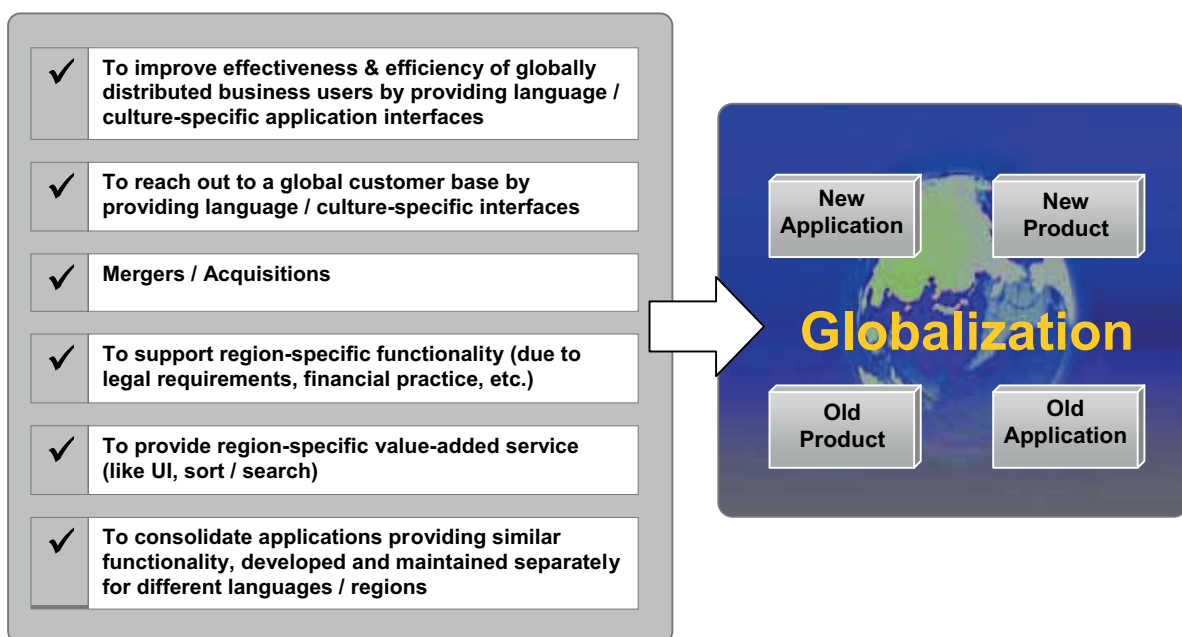
Introduction

Globalization and virtualization are changing the way enterprises do business and engage with customers, vendors, suppliers and all other business interfaces. To succeed in today's challenging global environment, enterprises must provide a local flavor to their business applications, services and products. This applies to existing applications and products, as well as those that are being developed. It essentially requires enterprises to address a wide range of linguistic, cultural, process, and technical challenges in their global applications and products.

Globalization is an integral part of business and IT. Enterprises must have an overall globalization vision and strategy to ensure the success of globalization programs. Application globalization can be achieved by following a methodology involving two high-level phases:

- Internationalization (i18n)
- Localization (l10n)

Globalization: Business Context



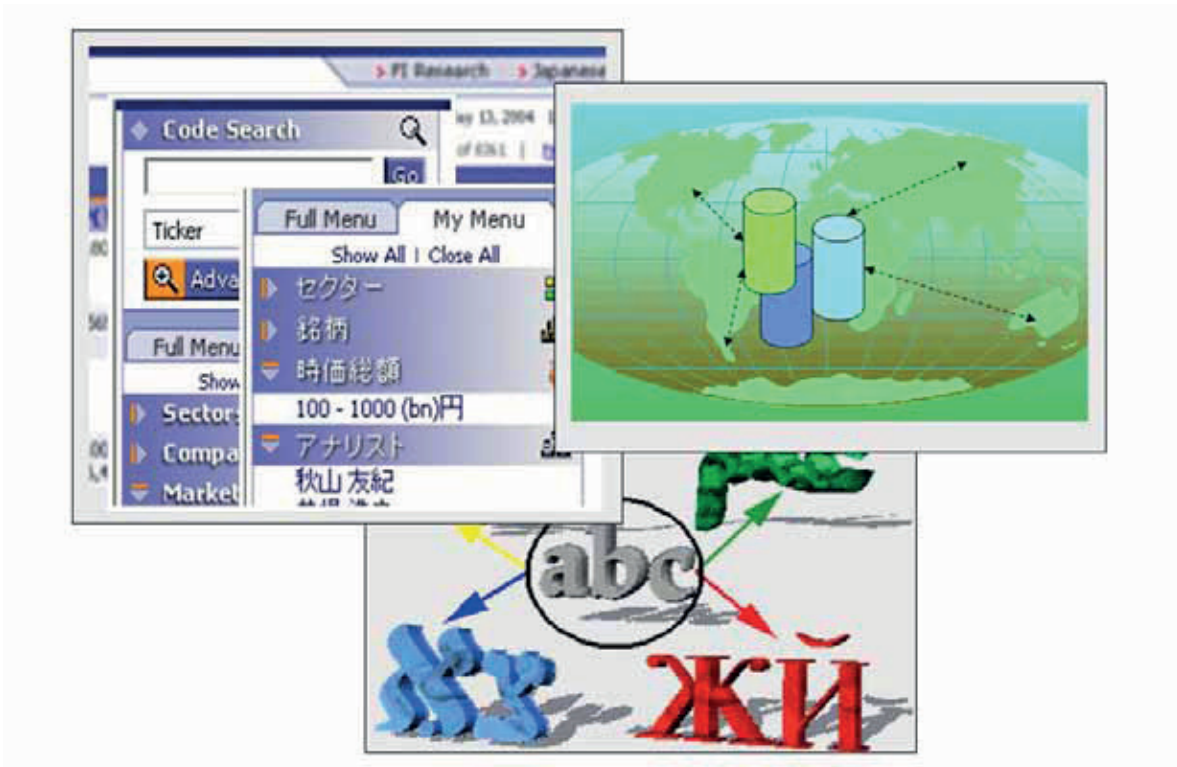
Enterprises are globalizing their applications for a variety of reasons. For organizations with a multi-cultural, multi-lingual workforce distributed across the globe, addressing local needs in applications promises improved effectiveness and productivity. For organizations marketing products across various countries, country- / region-specific functional and non-functional requirements – local laws / regulations, financial practices, industry-specific business rules, etc. are major concerns during a globalization effort. Often, organizations globalize their operations to reduce costs by unifying all applications providing similar functionality, but which are developed and maintained separately for language and cultural reasons. Globalization may also arise due to mergers and acquisitions.

Globalization Requirements

Globalization requirements for industry verticals emerge from business drivers and context. At the highest level, they can be classified as globalization feature requirements and globalization Quality of Service (QoS) requirements.

Globalization features

Globalization feature requirements address the language, cultural and functional requirements of a region or country without which an application / product cannot be used effectively. It can be further categorized as user interface requirements, data storage / retrieval requirements, and data processing requirements.



User Interface (UI) requirements:

- Display UI labels, status messages, error messages, etc. in the end-user's language
- Date format, number format, currencies, phone / fax numbers, addresses, units of measurement, time zones conform to the end user's region/country
- Cultural aspects like color and images are specific to the end-user's region / country

Data Storage / Retrieval requirements:

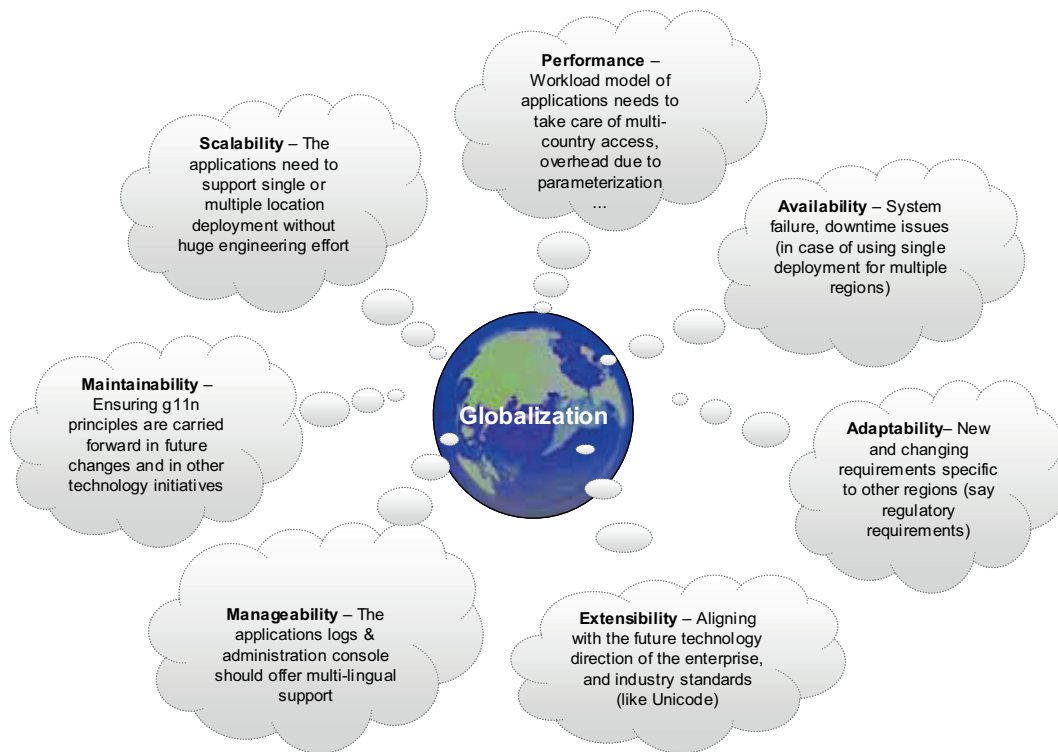
- Support for storage and retrieval of language-specific transaction data input by users in different languages without data corruption or loss
- Support for displaying and capturing language-specific reference data

Data Processing requirements:

- Support for different types of string sorting and comparison logic
- Support for interfacing with internal and external applications
- Support for locale specific algorithms, business rules, regulations, and processes

Globalization QoS

Though it is commonly perceived that identification of globalization features required in a particular application is the only requirement of globalization, QoS aspects also play a significant role in globalization of applications and should be captured with due diligence.



Some typical QoS aspects to be addressed while globalizing applications:

Performance – Globalization essentially means parameterization of applications, and designing applications for a large user base with different usage patterns across countries. Changes in the applications' performance behaviour are expected after globalization. All performance requirements aren't necessarily addressed even if the required globalization features are built into the application.

Scalability – Global applications need multiple deployment options (in central or multiple data centre locations in a single or multiple server model). They should be able to scale out and / or scale up without code changes.

Availability – Availability requirements are complex as users from different regions in different time zones may access the application simultaneously. Therefore, system failure, downtime planning, etc. have to be considered thoroughly at the design stage.

Manageability – When applications are deployed in multiple regions, local maintenance teams must be given support for analysis and administration in local languages. All system documents (installation script, log files, etc) need to be made available in the local language and format.

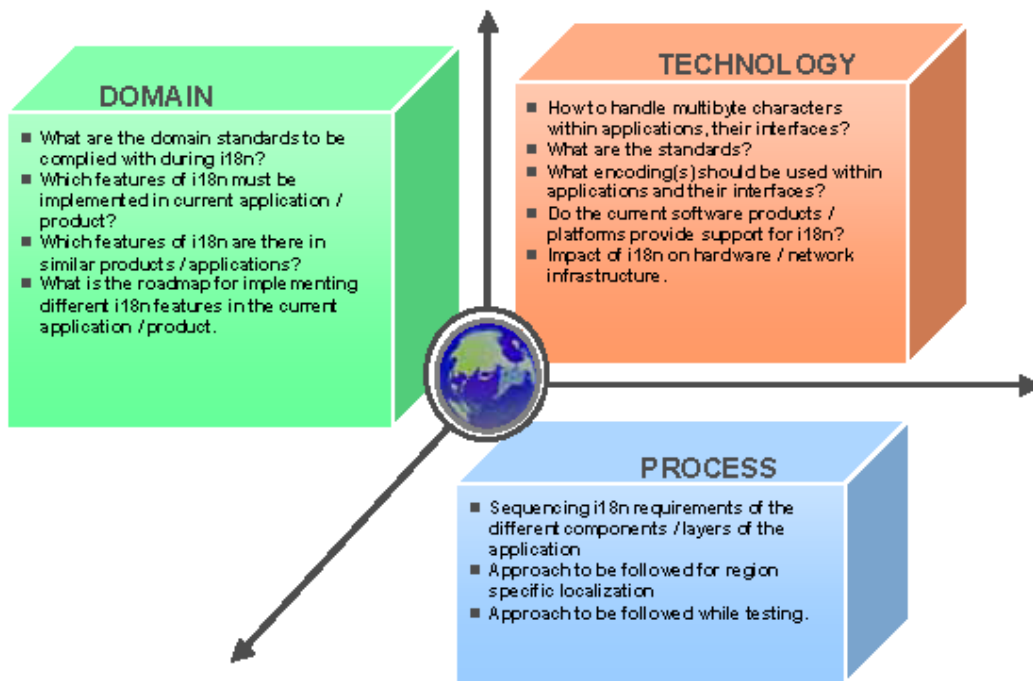
Extensibility – Application architecture and design must be aligned with industry standards and trends to ensure that the investment in globalization yield returns even with advancements in globalization technology. It must also be aligned with the organization's technology and business roadmap.

Adaptability – Architecture and design have to be adaptive to allow for development of new location-specific requirements without considerable change and in the shortest possible time span.

Maintainability – The principles and design guidelines of globalization should be considered while enhancing an application. A process has to be in place to ensure that minimal code changes are required to cater to the needs of different regions.

Globalization Challenges

High-level business drivers, together with specific globalization features and QoS requirements, contribute to the globalization challenges. These challenges are loosely categorized and discussed below.



Technology Challenges

Technical challenges are observed in the following categories:

Encoding and Coded Character Set related challenges –

- Unicode support and implementation mismatch between the products / platform:** These days applications are developed using multiple platforms, products and packages. Though Unicode is a standard supported by most of the current technical platforms / products / packages, the level of support and the way it is supported by different technical platforms can differ significantly (e.g. the type of Unicode Transformation Format). This poses a major challenge from the integration perspective.
- Use of language specific encoding in interfacing with existing products/application:** Though Unicode is the standard, applications often need to support language-specific legacy encodings (like Shift-JIS, EUC-JP for Japanese) for interfacing with other external or internal applications. It is sometimes difficult to ensure that the technical products / platforms used to build the application provide the required support for language-specific legacy encoding.
- Configuring encoding for different product / platform:** Configuring the right encoding at different product / platform levels is very challenging. This is due to the lack of documentation and explicit focus on globalization at the product / platform level. Improper or unavailable configuration at the platform / product level will limit options to using default OS level encoding. This can result in the application behaving waywardly or in data corruption / loss.
- Troubleshooting at the interface level:** Interfaces in an application are prone to data corruption / data loss due to mismatch of encoding. The byte sequence of multi-byte characters must be tracked to identify problems at the interface level. In most cases, this is difficult due to the unavailability of tools for byte level tracking at the interfaces.
- Supporting all characters for a given language:** Constraints in the underlying platform / technology to handle a large number of characters and the obsolescence of some characters may cause difficulties in supporting all characters of languages like Chinese, Japanese, Korean, etc.

Locale and Parameterization related challenges –

- *Lack of a standard definition of locale across platforms:* Platforms (like Windows .NET, Java, UNIX, OS400, etc.) have different definitions of locale. This creates problems in locale implementation. Using multiple platforms / products to develop an application leads to further complications. The seamless transfer of locale object across interfaces on different platforms / products is restricted by the lack of a standard definition.
- *Need for the appropriate tool for impact analysis and mass code change:* For globalization of an existing application, the existing code needs to be changed for parameterization of strings, images, number format, currency format, etc. (basically, for implementing globalization features selected for the application). For a huge code base and for applications written in older technologies like C, RPG, COBOL, etc., identification of the impacted areas and making changes is a daunting task. Not many tools are available for doing this and the time needed to make such changes is considerable.
- *Having a single code base:* It is sometimes difficult to produce a single code base when the region- / country-specific features are addressed through parameterization. This depends on the type of globalization features needed, as well as technical platform / products used in the applications.
- *Choice of Locale/Language:* Although desired by users, it may not be technically feasible to automate the locale/ language choice of applications.
- *Managing availability across all regions / time zones –* It is often difficult to arrive at a downtime schedule for maintenance of applications accessed from different regions/countries across the globe.

Presentation, Processing and Storage related challenges –

- *UI design:* It is very difficult to design UI screens that can accommodate string elements in any language, without jeopardizing the screen's overall layout because the equivalent of a 4- letter word in English may be 8 letters in German and may be expressed by 1 ideographic character in Japanese.
- *Non-uniformity of column definition for multi-byte characters across different database product:* The way in which databases store and process multi-byte characters varies considerably from one database product to another. The optimal database design (in terms of column length and type) has to be based on the database product used. This requires a sound understanding and insight into how a database product handles multi-byte characters.
- *Handling collation:* European languages, other than English, have certain characteristics like One-to-Two character mappings, Two-to-One character mappings, Diacritics, etc. These aspects have to be carefully taken considered while sorting the characters in those languages. Similarly, for ideographs in East Asian languages, sorting can happen in a number of ways, like stroke count based sorting, main radical based sorting, pronunciation based sorting, etc., and users should be able to select the sort of his choice. Most of the technical platforms / products provide support for achieving the required sorting logic / algorithm for non-English characters. However, not all types of sorting can be achieved and configuring the products/platform to achieve it is usually a complex task.
- *Migration of existing data:* Very often, databases contain multi-byte characters in languagespecific legacy encoding. Though moving the data to Unicode is the best practice, doing it can be challenging and risky.

Domain Challenges

Every industry vertical comes with its own set of globalization requirements and challenges. This is essentially because the business practice, even in a vertical, can vary from country to country. The challenges typically encountered at domain level are:

- *What are the domain standards and regulations to be complied with –* This aspect is often unclear and it needs due focus to identify the same for different countries as well as the internationally accepted standards for the application(s) in question.
- *Which features of g11n are there in similar products/applications in an industry vertical –* Similar products / applications from different companies may have some specific globalization features but not others. However, for the applications in question, the essential and optional ones must be identified carefully as implementing all desired globalization features may not be feasible within the stipulated timeline and cost. However, creating a roadmap for them considering all constraints is typically a challenge.

Process Challenges

Globalization programs typically face challenges in terms of managing the number of activities, stakeholders, timelines, dependencies and constraints. A robust, flexible and well thought-out process is a must. The typical process-level challenges in executing a Globalization program are:

- **Time to market** – It is always challenging to arrive at the right extent of globalization features needed to be added in the right set of applications for the target markets, given the tight deadlines.
- **Sequencing the activities and resolving dependencies** – Often, a globalization program means globalization of a set of applications. However, it also means identifying the impact on other applications, understanding the impact on other initiatives, getting support from the product/ platform vendors/ service vendors/ translation agencies and a comprehensive rollout plan in target countries/regions. Having a program plan with detailed activities to manage these aspects is an obvious challenge.
- **Ensuring Comprehensive Testing** - It is difficult to ensure comprehensive testing of a global application covering all required parameterization (based on the selected globalization features). Essentially, it extends the time required for the program. Further, testing for different languages/regions calls for local resources with language and cultural expertise. Performance testing is crucial, as a significant amount of parameterization is introduced in the application due to globalization.
- **Translation** – Translation during a globalization program can prove challenging. The troublesome translation aspects include cultural nuances, the format in which strings for translations are communicated, getting the right translation agency with expertise in all target languages, etc.
- **Continuity of i18n features** – Often, after the completion of the globalization program, the decision and guidelines are not carried forward for future changes/enhancements in the applications.
- **Lack of awareness about globalization in different stakeholders** – Linguistic and cultural differences, non-English characters, bi-directional scripts, globalization-related technology standards, globalization supports from technical products/platforms... building overall awareness in stakeholders of these aspects and subsequent complexities is a necessary but demanding challenge.

Conclusion

Any globalization program, irrespective of domain, technology or the number of applications at stake, throws up challenges in different dimensions rooted in globalization requirements or driven by the basic business objectives for globalization. Understanding specific globalization challenges in a program context is essential. It is important to address these for successful program execution. The number and complexity of challenges can be overwhelming. However, recognizing them and identifying key measures and decisions to be taken at different milestones can help manage the program effectively. A strategy definition phase, aimed at achieving the above objectives, is recommended and can prove rewarding in the long term.

Glossary

- Character Encoding - The mapping from a coded character set to a sequence of bits. Also known as encoding scheme.
- Coded Character Set - A collection of characters in which each character is assigned a unique numerical value.
- Internationalization (i18n) - The process of designing an application/product so that it can be adapted to various locales / regions without engineering changes.
- Locale - The encapsulation of language, formatting rules and cultural nuances for a specific region of the world
- Localization (l10n) - The process of adapting an internationalized software for a specific locale / region
- Multi-byte Characters - Some characters for languages like Japanese, Korean, Chinese need to be represented in more than one byte unlike US English (one character -> one byte) as the number of distinct symbols used in respective languages are more than 256
- QoS – Quality Of Service or Quality Requirements. (<http://www.sei.cmu.edu/community/rhasworkshop/azuma.pdf#search=%22ISO%2FIEC%209126%20%22>)
- Unicode - A fixed-width 16-bit character encoding used to encode the world's writing systems. Unicode is maintained by the Unicode Consortium (<http://www.unicode.org/>)
- Unicode Transformation Format - This is a generic name for several encoding schemes supported by Unicode Standard. (http://www.unicode.org/faq/utf_bom.html#1)

About the Author

Sourav Mazumder is a Principal Architect at Infosys, with 10 years of experience in the IT industry. He has worked on projects dealing with internationalization and multi-byte characters, especially dealing with East Asian characters. He has also worked in different phases of typical i18n projects, from byte-level trouble shooting to technical consultancy to clients in their i18n initiatives.

Sourav is an iCMG certified Software Architect as well as TOGAF 8 certified practitioner.



For more information, contact askus@infosys.com

About Infosys

Many of the world's most successful organizations rely on Infosys to deliver measurable business value. Infosys provides business consulting, technology, engineering and outsourcing services to help clients in over 30 countries build tomorrow's enterprise.

For more information about Infosys (NASDAQ:INFY), visit www.infosys.com.