

## **REDUCING LOCALIZATION COSTS WITH XML-BASED TECHNOLOGY**

Daniel W. Dube  
Director, Business Development  
ISOGEN International

### **Overview**

This paper is intended to build on papers previously presented at ASLIB *Translating and the Computer 21* and *22*. We will summarize how XML-based technology has enabled some Global 2000 companies to streamline their localization processes and reduce the cost of delivering localized information products, and we will document the significant progress that has been made in the last two years in the area of standards-based technology.

In addition, we will present business case studies based on actual implementation projects. We will examine the methodologies used to help these customers migrate to an XML-based production environment, the issues that were discussed and addressed during the migration effort, and the real business results that these companies enjoyed as a result of implementing XML-based tools into their departments.

### **"Typical" Localization Process**

Today, most companies with a need to deliver information to a global audience still rely on a manual, paper-based process for localizing their content. Human intervention is required at all stages, from the origination of source language data through review, quality assurance, and final delivery of localized information. In most cases, authors are working with tools such as MS-Word and FrameMaker and sending complete document instances to a translation vendor. Because these files are stored and managed at the document level, authors have no easy way to identify only the pieces of content that have changed between revision cycles. As a result, authors continue to send entire documents to the translation vendor for localization, even if only a small portion of the content has been changed.

There are inherent inefficiencies to this process, which lead to the following typical results:

- " Longer translation turnaround time
- Higher localization costs
- Lost revenue opportunities

### **XML-Based Technology Components**

Over the last five years, technologies based on XML have evolved to meet the needs of information producers and suppliers. We have leveraged and enhanced some of these commercial products to streamline the localization process. In general, there are three areas of functionality that have been proven to provide strong results:

- XML-based content management/workflow systems that leverage the reuse and repurposing of XML information components

- "Knowledge Broker" portal environment to manage linking of content from disparate applications
- XSLFO technology to automate the rendering of multilingual content for print delivery

The following sections will provide more details on each of these functional areas of XML-based technology.

## 1. Content Management/Workflow Systems

In an XML-based content management production environment, information is managed at a lower level of granularity than a document. This allows for the *reuse of information at the component level*. For example, many companies standardize on boilerplate content, such as warnings and cautions, which are approved by corporate legal departments. In an XML-based content management environment, these warnings and cautions are written and stored once in the repository, and are shared by any documents that need to include these warnings and cautions. If the content of a warning needs to be updated, it only needs to be edited in one place, and all occurrences of that warning will immediately be updated to reflect the latest version of the approved content. This also means that components only need to be translated once, and all target language documents that contain this content will also have the latest localized version available.

The more robust systems in this class of technology also have integrated workflow tools. This functionality can be extremely effective for automating many of the manual tasks associated with producing source language and localized content products. For example, an automated workflow task can be configured to automatically build a "localization package" that contains all source language content objects in the repository that have changed since the last translation cycle, and automatically export them to the localization vendor. (No human intervention required!)

Consider the following case study for an example of the power of this type of functionality:

### **Case Study #1: Multinational High Technology Manufacturer - Consumer Products**

#### ***Situation***

A technical publications group for this Global 2000 company is responsible for providing hardcopy documentation, web-based help, and integrated online help for their product lines. Their global customer base required that this information be delivered simultaneously to all of their markets in 33 languages. Due to the volume of information produced, this company outsourced the localization and composition of the localized information products to several of the largest localization service providers. Company management was continuing to plan the rollout of new products, while at the same time expecting to maintain or reduce staff headcount in the technical publications department.

#### ***Critical Issues***

- The production process was paper-driven and highly manual in nature, which made project-related information difficult to manage and track.
- The group was unable to share and reuse information due to a mix of proprietary authoring and publishing tools (e.g., FrameMaker, MS-Word, Robohelp, etc.). This led to much "cutting and pasting" of content between documents.

- "Heroic" actions were required by the staff in order to meet the manufacturing release dates. Late delivery was simply not an option, so employees were required to work many nights and weekends as deadlines approached.
- Staff members performed many tasks that were redundant or unnecessary within these tight timeframes, which increased frustration and stress. For example, time was spent "undoing" formatting done in FrameMaker for print output in order to prepare the same content for online help delivery.
- In general, the localization process was too time-consuming, too expensive, and too inflexible.

### **Solution**

- Migrated the customer from a proprietary authoring/publishing environment to a standards-based infrastructure of reusable XML information components
- Implemented an XML-based content management/workflow system to leverage reuse of content, version control, automatic link validation, and advanced search capabilities
- Set up workflow functionality to automate and streamline the localization process

### **Results**

- The system was implemented and in full production within 8 months (on time and on budget). The implementation was spread across two phases:
  - Phase 1: Configuration of English-language XML environment (4 months)
  - Phase 2: Configuration of multilingual content and workflow with localization vendor (4 months)
- The entire cost of the system and implementation services was recovered after the first translation cycle! This was primarily due to the reduced translation and desktop publishing costs associated with translation of XML components, and with the more efficient process in place that eliminated the redundant efforts and content.
- This department now has significantly more flexibility to support an increasing workload with the same staff headcount.

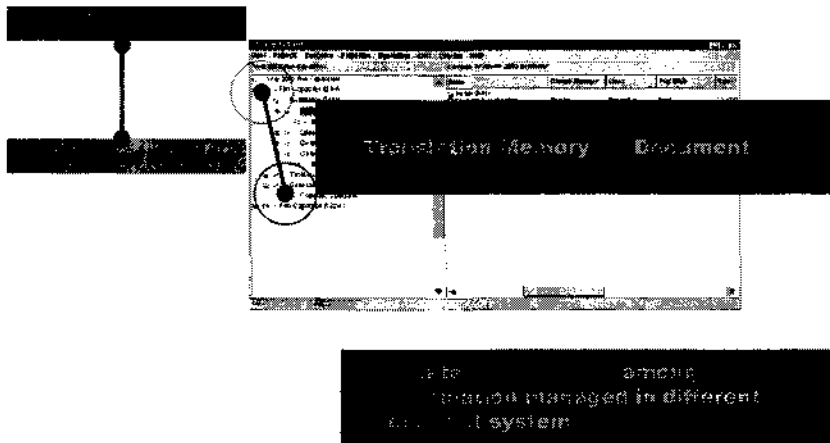
## **2. The Next Wave: Knowledge Brokers**

In today's world, the localization process involves collaboration of information stored in many related, yet disparate, applications, including:

- Content management systems (e.g., Documentum, Interwoven, Astoria)
- Translation memory tools (e.g., Trados)
- Terminology databases and glossaries
- Digital Asset Management (DAM) systems

"Knowledge Broker" is our code name for XML-based portal technology that shows the promise of being able to provide content owners with the ability to link information stored in these distributed, heterogeneous environments. A portal can act as a client to retrieve resources and provide a virtual view of a collection of information. The goals of this technology will be to:





### Streamline the Localization Process with Knowledge Brokers

Once you have linked relevant information objects and created associations between them with appropriate business rules, it is possible to further automate the localization process by assigning operations to linked information collections. The following diagram illustrates this concept:



## **Proven Results**

We can point to some examples of companies that have implemented this type of solution:

### ***Major Engine Manufacturer***

A global engine manufacturer has implemented a system that leverages linking between XML and their internally managed Trados TM. Some statistical results from this project include:

- Currently have an automated localization workflow that supports delivery of information products in six languages
- Reduced production time of a localized 400-page technical manual from 8 months to 10 weeks
- Reduced translation costs by more than 70%, primarily due to reuse of XML information components, link management between XML content and the Trados TM, and an automated composition system to lay out localized manuals without human intervention

### ***Tier 1 Supplier to Automotive Industry***

A supplier to many of the world's leading automotive manufacturers successfully implemented a content management/publishing system to outsource the creation of vehicle documentation for worldwide distribution. Some statistical results from this project include:

- Currently have an automated localization workflow that supports delivery of information in over 30 languages
- 90% of the XML information components are shared between two or more vehicles
- Turnaround time to deliver localized information products for all required markets was reduced from 6 months to 2 weeks

## **3. The Newest Breakthrough: XSLFO Rendering Tools**

The technology discussed so far has dealt with creation and localization of content from the aspect of managing content in a more efficient way. However, there is still one area where there has been room for improvement: the desktop publishing (DTP) of localized content. According to customers we have discussed this topic with, DTP remains the most expensive service component associated with localization. This is due to the fact that many languages require a manual layout of content with a DTP tool, such as FrameMaker. Examples of these types of languages include:

- Double-byte character sets (e.g., Asian languages)
- Languages with complicated hyphenation routines (e.g., Thai)
- Right-to-Left (RTL) languages (e.g., Hebrew, Arabic)

Until recently, there has not been technology available to automate the layout of languages with these difficult formatting issues.

However, last year the W3C body issued a set of recommendations to supplement the XSL language for expressing stylesheets. Many people are already familiar with the XSLT recommendations for transforming XML documents (W3C Recommendation 16 November 1999). This has been an effective mechanism for rendering XML documents for HTML-based browser viewing.

The newest recommendation is in support of XSL Formatting Objects (XSLFO). XSLFO presents an XML vocabulary for specifying formatting semantics (W3C Recommendation 15 October 2001). XSLFO provides a framework to create a tree of formatting objects to define composition rules for XML documents down to the lowest level elements. The resulting "FO file" can be interpreted by an XSLFO-aware composition engine to automatically render print-ready document formats, such as PDF or PostScript. We have successfully utilized technology based on XSLFO to automate the rendering of print-ready content into virtually all languages.

## **Case Study #2: Multinational High Technology Manufacturer - Business Products**

### ***Situation***

A technical publications group for this Global 2000 company is responsible for providing hardcopy documentation and web-based help for their product lines. Their global customer base required that this information be delivered simultaneously to all of their markets in 34 languages. Due to the volume of information produced, this company outsourced the localization and composition of the localized information products to several of the largest localization service providers. Company management was continuing to plan the rollout of new products, while at the same time expecting to maintain or reduce staff headcount in the technical publications department.

This department had already realized some cost savings associated with reuse of XML content; they had migrated to an XML-based authoring and publishing environment based on the Arbortext Epic product suite.

### ***Critical Issues***

- Although localization costs had been reduced due to migration to XML, there was still a high cost associated with DTP of localized information
- The Epic product does not support composition of all the languages required by this customer; therefore, there was still a need to hire out DTP services for a subset of languages
- The turnaround time for DTP of complex languages was too long
- Manual layout of content led to consistency issues in the presentation of content
- Redundant efforts were still required to produce both online and published hardcopy versions of content for languages that utilized DTP

### ***Solution***

- Implemented an XSLFO-compliant formatting system (XSL Formatter, from Antenna House) to supplement the existing Arbortext Epic environment
- Utilized the ISOGEN 118N toolkit to enhance the XSL Formatter to provide native support for all required languages
- Added rules-based intelligence to automatically handle complex multilingual layout tasks:
  - o Hyphenation for complex languages (Thai)
  - o Automated layout of RTL languages, including the insertion of LTR text within a RTL sentence (e.g., insertion of an English phrase within an Arabic sentence)



- Created one stylesheet that drives the layout for all languages in both print and online deliverables

### **Results**

- This environment was fully implemented within 12 weeks
- The new XSLFO environment was approximately 60% faster to develop than the original FOSI-based environment, and it was able to address the language limitations of the FOSI environment
- The entire cost of the tools and implementation services was recouped in the first production run, due to the fact that DTP services from the localization vendor were no longer required

## **Conclusions**

Standards-based technologies (primarily based on the XML family of recommendations from W3C) have continued to evolve and mature. As more companies have adopted the use of applications based on these standards, creative use of these tools has resulted in dramatic cost savings related to localization and publishing of localized products. This is now a low-risk and smart area of investment for companies that have a need to deliver localized information to a global audience.