

Digital content for global audiences

These days the production of digital content runs at high speed placing enormous pressure on global organizations to localize it. Digital content can be classified as enterprise content and personal content. We will examine each of these and take a close look at the challenges faced by enterprises localizing digital content. We will also offer a potential solution to one of the most significant problems encountered: the cost of localization.

Digital Content

Digital content is data stored or published on electronic devices such as computers, mobile phones, televisions and digital cameras. It usually consists of text, audio, video, images, and software. We are living in the midst of a digital revolution that started in the 1970s. Since the widespread adoption of the home computer in the 1990s, many people are now able to share photos online, retrieve information via the web and communicate via email or short message service (SMS).

Information from almost every field – including business, law, medicine, science and religion – is now published as digital content. Ferreras (2009) states “the world of scholarly communications is deeply affected by the on-going change of medium...several academic objects of prime importance, such as Master’s theses, doctoral dissertations and refereed publications need to be re-defined within the digital landscape.” We can conclude, therefore, that almost anything that exists in the physical world will have a presence on the web. We realize that there are many languages, which do not have much penetration in the web, but this topic is beyond the scope of this article.

We will now focus on commercial enterprises and how they develop multilingual digital content. Many organizations localize digital content for users in different locales, usually to increase international sales revenues. Generating multilingual versions of enterprise content requires significant investment in people, processes and technologies.

Localization professionals include translators, proofreaders, engineers, software testers and project managers working for software vendors, tool developers and service providers. Localization processes such as translation, engineering, quality assurance and project management require internal coordination or outsourcing to language vendors. Tools providers publish and sell the following three types of localization technologies:

1. Computer-assisted translation (CAT) tools: Translation memory (TM) tools, terminology management tools, alignment tools and tag editors.
2. Machine translation (MT) systems: Rule-based MT tools, statistical-based MT tools, example-based MT tools and hybrid MT tools.

3. Software localization tools: Localization suites, internationalization tools, engineering tools, quality assurance (QA) and testing tools, and project management (PM) tools.

The total investment required in people, processes and technologies depends on the scope of the localization project in terms of the type of content being localized as well as the volume and number of languages involved.

In a world where digital content and globalization are growing rapidly, many organizations invest significant human and monetary resources to test, engineer and edit localized versions of this content.

Enterprise content

Enterprise content is usually high-volume and static, developed by professionals, and published for commercial reasons. It includes commercial software, corporate websites, e-Books, marketing communications and so on. Enterprise content management (ECM) refers to the technologies, strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to an organization and its processes¹.

Personal content

Personal content is developed by non-professional users for social reasons. It is low-volume, dynamic and non-repetitive. Personal content is usually located in forums, blogs, emails, online gaming communities and SMS messages. Personal content management tools help people organize their personal information, and to define who is allowed to access it.² They often provide features such as download areas, blogs, forums and image galleries.

Challenges

One of the biggest challenges facing digital content publishers is how to reduce the cost of localizing digital content without sacrificing its quality. Several strategies are employed by organizations to reduce the cost of localization, including outsourcing tasks to low cost vendors, using MT systems to automate localization processes, and utilizing crowdsourcing or community-based platforms to encourage users to volunteer translation services.

Cost saving through internationalization

Let's focus on another cost-reducing strategy: internationalization. The Localization Industry Standards Association (LISA) defines internationalization as "the process of enabling digital content for localization at a technical level"³. It involves authoring digital content for global audiences, and engineering it so that no cosmetic or functionality issues occur during translation. Therefore, internationalization is implemented before translation in order to detect and resolve any problems as early as possible in the localization process. Enterprises can save money if they internationalize a product in advance. Otherwise they might face additional expenditure during post-translation testing and engineering. LISA estimates that "it

takes twice as long and costs twice as much to localize a product if it is not properly internationalized to start with.”

The main challenge when developing digital content for global audiences is to ensure that no linguistic, cultural and technological issues will occur during translation. Linguistic issues such as poor grammar or unclear syntax may be addressed with controlled language, style guides, terminology databases and so on. Cultural issues such as inappropriate references to sensitive religious or political topics can be avoided with cultural research. Technical issues like incorrect character encoding may be alleviated with pre-translation software testing. These issues can be avoided by building a web application that facilitates checking the linguistic, cultural and technical integrity of source language digital content. This results in digital content that is easier to translate for both human translators and machine translation systems.

Implementing internationalization using a web application

Digital content needs to be internationalized due to the diversity that exists in different locales:

1. Linguistic diversity: Different languages in different locales; even the same language spoken in different regions may have different spellings, grammar and punctuation;
2. Cultural diversity: Different religions, symbols, political systems, color associations, customs and so on in different locales.
3. Technological diversity: Different character encodings, computer hardware and internet browsers in different locales.

The main objectives of the application are:

- to enable professional content developers to publish source language digital content that is more usable for local users and translatable for human translators and MT systems;
- to enable localization professionals to assess the translatability of the source language files they receive from professional digital content publishers;
- to enable localization project managers to scope the resources required to localize the source language files they receive from professional digital content publishers.

The users of the system will, therefore, include content developers, localization professionals, managers, and academics.

The following three features of a web application help to address the challenges of internationalizing digital content: a test area, a digital library and a virtual community. The test area enables the user to upload files to the system, and run a number of specified QA checks. It also enables users to view project statistics, generate reports and pseudo-translate files. Whereas the test area contains most of the functionality of the system, the digital library will store a vast repository of guidelines relevant to the production of internationalized digital content. These guidelines will include content development guidelines, cultural research and industry standards. The final

component of the system is the virtual community, where users will access forums, upload and download resources, and connect with other users.

The benefit of this application is that it offers content developers a functional Web 2.0 interface with which to check how internationalized and translatable their digital content is. In addition, it also provides a repository of useful guidelines, downloadable resources and a virtual community with which ideas and opinions can be shared with other system users.