Handling Multilingual Content in Digital Media: The Multilingual Information Framework

Samuel Cruz-Lara, Satyendra Kumar Gupta, Laurent Romary

HANDLING MULTILINGUAL CONTENT IN DIGITAL MEDIA: THE MULTILINGUAL INFORMATION FRAMEWORK

S. CRUZ-LARA, S. GUPTA AND L. ROMARY

LORIA (UMR 7503) / INRIA Lorraine
Campus Scientifique – BP 239
54506 Vandœuvre-lès-Nancy, FRANCE
{cruzlara, gupta, romary}@loria.fr

This document presents MLIF (Multi Lingual Information Framework), a high-level model for describing multilingual data across a wide range of possible applications in the translation/localization process within several multimedia domains (e.g. broadcasting interactive programs within a multilingual community). There are at present numerous different standards available for dealing with translation and localization activities. All these formats work well in the specific field they are designed for, but they lack a synergy that would make them interoperable when using one type of information in a slightly different context, giving rise to the fear of competition between them. MLIF defines a generic model for all these formats, from which one can derive precise specifications for each of them.

1. Introduction

Linguistic information plays an essential role in the management of multimedia information, as it bears most of the descriptive content associated with more visual information. Depending on the context, it may be seen as the primary content (text illustrated by pictures or videos), as documentary content for multimedia information, or as one among several possible information components in specific contexts such as interactive multimedia applications. Linguistic information can appear in various formats: spoken data in an audio or video sequence, implicit data appearing on an image (caption, tags, etc.) or textual information that may be further presented to the user graphically or via a text to speech processor.

In this context, dealing with multilingual information is crucial to adapting the content to specific user targets. It requires one to consider potential situations where the linguistic information contained in a multimedia sequence is either already conceived in such way that it can be adapted on the fly to the linguistic needs of user, or by using an additional process where content should be adapted before presenting it to the user.

Finally, there are a wide variety of applications within which multilingual information may appear, which supports development of a generic framework

for dealing with multilingual content: subtitling of video content, dialogue prompts, menus in interactive TV, descriptive information for multimedia scenes, karaoke management, etc. Such information should be considered in the light of the experience of more specialized communities traditionally dealing with multilingual content, namely the translation and localization industry.

2. Background

The scope of research and development in localization and translation memory (TM) process development is very large. There are numerous independent groups LISA [1], OASIS [2], W3C [3], ISO [4] working on these aspects namely LISA, OASIS, W3C, ISO, etc. LISA is working for the GILT (Globalization, Internationalization, Localization and Translation) business community. It has evolved as the premiere organization for developing language-technology standards. OSCAR [5] "Open Standards for Container/Content Allowing Re-use" is LISA's special interest group (SIG) for the creation of open standards. OASIS evolved to drive the development, convergence, and adoption of structured information standards in the areas of e-business, web services, etc. OASIS is driven by various technical committees (TC) formed by its members. OASIS XML Localization Interchange File Format TC (OASIS XLIFF TC) was formed with the purpose to define, through XML vocabularies, an extensible specification for interchange of localization information. The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to bring the Web to its fullest potential. W3C organizes the work necessary for the development or evolution of Web technologies into Activities. Most of the activities under W3C are developing specifications/tools for management of language/text on the web. ISO is a network of national standards institutes from 148 countries working in partnership with international organizations, governments, industry and business and consumer representatives. ISO is a bridge between public and private sectors. It has also formed some technical committees, working in language-technology and IT applications in information, documentation and publishing.

Under the guidance of above-mentioned groups, many formats have been developed. Some of the major formats of specific interest for localization and TM are TBX [6], TMX [7] (LISA/OSCAR), XLIFF [8] (OASIS), Timed Text [9] (W3C), TMF [10] (ISO). TBX is an open XML-based standard format for terminological data. This capability will greatly facilitate the flow of terminological information throughout the information cycle both inside an organization and with outside service providers. TMX is a vendor-neutral, open standard for storing and exchanging translation memories created by Computer

Aided Translation (CAT) and localization tools. The purpose of TMX is to allow easier exchange of translation memory data between tools and/or translation vendors with little or no loss of critical data during the process. The purpose of XLIFF vocabulary is to store localizable data and carry it from one step of the localization process to the other, while allowing interoperability between tools. The Timed-Text specification covers real time subtitling of foreign-language movies, captioning for people lacking audio devices or having hearing impairments, karaoke, scrolling news items or teleprompter applications. It provides interoperability between different existing formats. ISO 16642:2003 specifies a framework designed to provide guidance on the basic principles for representing data recorded in terminological data collections. This framework includes a meta-model and methods for describing specific terminological mark-up languages (TMLs) expressed in XML. The mechanisms for implementing constraints in a TML are also defined in ISO 16642:2003. ISO 16642:2003 is designed to support the development and use of computer applications for terminological data and the exchange of such data between different applications. In the framework of digital media, MPEG4 and MPEG7 deal with multilingual data. MPEG-4 is standard for multimedia for the fixed and mobile web and MPEG-7 is standard for description and search of audio and visual content.

When we closely examine the different standards or formats developed by these groups, we find that they have many overlapping features. For example all these formats are based on XML schemas, provide extensibility, and bridge the gap between two systems or tools in different languages. There are many identical requirements for all the formats irrespective of the differences in final output. For example, all the formats aim at being user-friendly, easy-to-learn, and at reusing existing databases or knowledge. All these formats work well in the specific field they are designed for, but they lack a synergy that would make them interoperable when using one type of information in a slightly different context, giving rise to the fear of competition between them.

3. The Multi Lingual Information Framework

The Multi Lingual Information Framework (MLIF) is designed with the objective of providing a common platform for all the existing tools developed by the groups listed in the previous section. It promotes the use of a common framework for the future development of several different formats: TBX, TMX, XLIFF, Timed Text, TMF, etc. It does not create a complete new format from scratch, but suggests that the overlapping issues should be handled independently and separately. It will save time and energy for different groups

and will provide synergy to work in collaboration. Presently, all the groups are working independently and do not have any mechanism for taking advantage of each other's tools. MLIF proposes to concentrate on only those specific issues that are different from others and specific to one format only, so it will create a smaller domain for the groups' developers. It gives more time to concentrate on a subset of the problems they are currently dealing with and creates a niche that helps in providing a better solution for problems of multilingual data handling and translation issues.

In MLIF, we deal with the issue of overlap between the existing formats. MLIF involves the development of an API through which all these formats will be integrated into the core MLIF structure. This is done through the identification and a selection of data categories as stated in ISO DIS 12620-1[1] (in ISO/TC 37/SC 3[2]). MLIF can be considered as a parent for all the formats[3] that we have mentioned before. Since all these formats deal with multilingual data expressed in the form of segments or text units they can all be stored, manipulated and translated in a similar manner. This kind of data can easily be stored in data categories and in terminological mark-up. The results of IST SALT project [11] clearly show that it is not difficult to edit, store and reuse data categories. The SALT project combines two interchange formats: OLIF [12], which focuses on the interchange of data among *lexbase* resources from various machine translation systems, and MARTIF [13], which facilitates the interchange of *termbase* resources with conceptual data models ranging from the simple to the sophisticated. It provides a graphical user interface that can be used to access or to define new data categories or modify them.

4. Implementing language translation by using MLIF: the ITEA "Jules Verne" project experience

4.1. *The RAMO Model*

The RAMO (Reactive & Adaptive Multimedia Objects) concept has been developed within the ITEA "Jules Verne" project [14]. The main objective of this project is to jointly exploit the capabilities of the interactive digital

[1] Computer applications in terminology -- Data categories -- Part 1: Model for description and procedures for maintenance of data category registries for language resources.
[2] Computer applications for terminology.
[3] Formats like "Timed Text" deal with monolingual data explicitly, but can easily be extended to handle multilingual data.

broadcast industry for content creation to those of future home terminals and networks. RAMO is based on the notion that a new dimension of interactivity can be achieved by enabling multimedia objects to fulfil the following criteria: become fully autonomous, be independent from predefined scenarios, and fully emulate the characteristics and behaviours of the represented entities.

Such objects are able to react and adapt themselves to any contextual situation resulting from interactions with other objects of the application and/or from user actions. Interactivity in such a concept has to be seen at two different levels: the object level and the system level (running environment). Interactivity is not deterministic at the system level. At the object level it deals with two kinds of scenarios: predefined and evolving ones. Predefined scenarios consist in predefined behaviours that are initially set up and based on a stimulus/reaction model. The evolving scenarios of an object refer to the notion of intelligent adaptive and learning entities. The dynamic evolution of the scene composition, the contextual situations and the objects adaptations create multiple and complex combinations that lead to unpredictable solutions. Mixed time-based and event-based multimedia scenarios will be designed.

The RAMO model will be based on a flexible standard able to describe a scene taking into account the following object configurations: structures, relations, constraints, and attributes.

In addition to this, it should specify the basis upon which object rendering will be performed for the corresponding contexts with the use of the relevant technologies in order to be platform independent.

A set of RAMO MPEG-7 based description schemes (RAMO DS) is being developed in order to enable to completely describe RAMO objects in terms of states and behaviours. The states will be described based on “Sensorial / AV Representations” and in terms of “Temporal and Spatial Matches”. The behaviours will be described in terms of “Processes” and of “Events”. The MPEG-7 Multimedia Description Schemes (MPEG7-MDS) is an interesting technical choice to describe the RAMO Scenes and Objects. It brings standardized means to the description and the specification of the following main features:

- The structures of entities and their internal relations between elements;
- The multimedia content for attributes with their type definition;
- The multimedia resources for control and access;
- The references to external resources;
- The relations and interactions among entities;
- The granularity, the flexibility and the openness that the XML standard offers.

The RAMO model has to specify the communication protocols set up between the objects as well as the way RAMO components (Applications, Scenes and Objects) are processed. This has to be achieved starting from the object management level up to the system level (application supervision). The RAMO model aims at being a generic model usable for many kinds of interactive multimedia presentations. It should be implementation independent and open to and compatible with extension profiles.

4.2. *RAMO and MLIF*

The main idea is to create specific Data Categories dedicated to generic RAMO scenes, objects and their related properties and to use MLIF as a meta-model [15], [16].

High-level categories may consist in RAMO Actor Data Categories, Behaviour Data Categories and Event Data Categories. Low-level categories of generic RAMO descriptors may be media-oriented descriptors, directly inspired by the MPEG-7 standard; for instance Video Data Categories, Audio Data Categories, Animation Scenario Data Categories, etc.

Currently, within the framework of ITEA “Jules Verne” project, we are experimenting with some basic scenarios by using XMT (“eXtensible MPEG4 Textual format”) [17] and SMIL (“Synchronized Multimedia Integration Language”) [18].

XMT has been designed to provide an exchangeable format between content authors while preserving the author's intentions in a high-level textual format. In addition to providing an author-friendly abstraction of the underlying MPEG-4 technologies, another important consideration for the XMT design was to respect existing practices of content authors such as the Web3D X3D, W3C SMIL and HTML. XMT is suitable for many uses including manually authored content as well as machine-generated content using multimedia database material and templates. XMT may be encoded and stored in the exchangeable mp4 binary file or may also be encoded directly into streams and transmitted. XMT encoding and delivery hints exist to assist this process.

The “Synchronized Multimedia Integration Language” enables simple authoring of interactive audiovisual presentations. SMIL is typically used for "rich media"/multimedia presentations which integrate streaming audio and video with images, text or any other media type. SMIL is an easy-to-learn HTML-like language, and many SMIL presentations are written using a simple text-editor. SMIL has the two following design goals:

- Define an XML-based language that allows authors to write interactive multimedia presentations. Using SMIL, an author can describe the temporal behaviour of a multimedia presentation, associate hyperlinks with media objects and describe the layout of the presentation on a screen.
- Allow reuse of SMIL syntax and semantics in other XML-based languages, in particular those which need to represent timing and synchronization. For example, SMIL 2.0 components are used for integrating timing into XHTML and into SVG.

The general idea is that we may want to obtain, by means of a translation / localization process, a document in English corresponding to an XMT or SMIL document which is in French. Figure 1 shows how the localization process is performed.

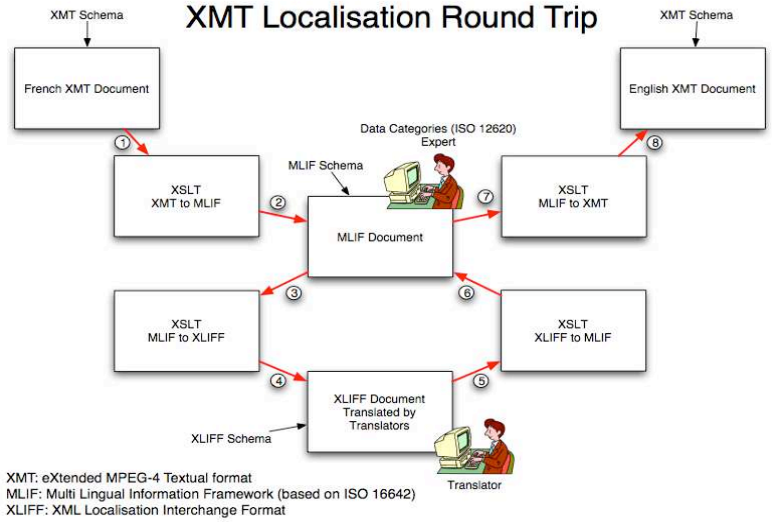

Figure 1. XMT Localization Roundtrip.

1. The original XMT_{french} document contains linguistic information in French.
2. Transformation of the XMT_{french} document into an $MLIF_{french}$ document.
3. Transformation of the $MLIF_{french}$ document into an $XLIFF_{french}$ document.
4. By using existing XLIFF environments, a professional translator performs French-English translation. We obtain an $XLIFF_{English}$ document.
5. / 6. Transformation of the $XLIFF_{English}$ document into an $MLIF_{English}$ document.
7. / 8. Transformation of the $MLIF_{English}$ document into an $XMT_{English}$ document.

5. Conclusion

We have presented MLIF (Multi Lingual Information Framework): a high-level model for describing multilingual data. MLIF can be used in a wide range of possible applications in the translation/localization process in several multimedia domains. In order to test MLIF in multimedia, we have proposed using MLIF and Data Categories for specifying RAMO (Reactive & Adaptive Multimedia Objects) within the ITEA "Jules Verne" Project. Currently, we are working on specifying the MLIF API through which a wide variety of formats (i.e. TBX, TMX, XLIFF, etc.) may be integrated into the core MLIF structure.

References and bibliography

1. LISA. www.lisa.org
2. OASIS. www.oasis-open.org/home/index.php
3. W3C. www.w3c.org
4. ISO. www.iso.org
5. OSCAR. www.lisa.org/oscar
6. TermBase eXchange Standard. www.lisa.org/tbx
7. Translation Memory eXchange Standard. www.lisa.org/tmx
8. XML Localization Interchange File Format. www.oasisopen.org/committees/tc_home.php?wg_abbrev=xliff
9. Timed Text. www.w3.org/AudioVideo/TT/
10. Terminological Markup Framework. www.loria.fr/projets/TMF/tmf.html
11. Standards-based Access to multilingual Lexicons and Terminologies www.loria.fr/projets/SALT
12. Open Lexicon Interchange Format. www.olif.net/
13. MAchine-Readable Terminology Interchange Format. ISO 12200:1999.
14. S. Renault, F. Meinköhn, D. Khadraoui, and P. Blandin. "Reactive and adaptative multimedia object approach for interactive and immersive applications". Proc. IEEE ICTTA 2004, Damascus, Syria, 2004.
15. S. Cruz-Lara, L. Romary, Ph. Sébire. "Using the concept of Data Category Registry for specifying RAMO structure". ITEA "Jules Verne" Project. Project Management Team. Luxembourg. November 2003.
16. S. Cruz-Lara, L. Romary, Ph. Sébire. "Managing Terminology and Multilingual Content within RAMO". ITEA "Jules Verne" Project. Project Management Team. Luxembourg. November 2003.
17. M. Kim & S. Wood. "XMT: MPEG-4 Textual Format for Cross-Standard Interoperability". IBM T.J Watson Research
18. Synchronized Multimedia Integration Language (SMIL 2.0) http://www.w3.org/TR/smil20/