Translation Process Metadata

Linked Open Data and MultilingualWeb-LT Requirements
Pedro L. Díez Orzas (Linguaserve)
Dublin, June 13th

Process

▶ These categories are used primarily for controlling or indicating the state of the content production process:

- ◦ **readiness**: localisation workflows and processes
- ◦ **progress indicator**
- ◦ **localisationCache**: real time localisation system has to cache the content or a part of the content for future

Readiness

ITS2.0 should be able to indicate:

- readiness for a particular process:
- priority
- expectation of when an a specific process should be completed
- if an element previously submitted to a process
- Other ??

Other "readiness" attributes?

- *targetlanguages* – value language ID
- *contentType* – values: MIME version: X.X
- *pivotLang* – value language ID
- *contentResultSource* –value: yes / no. Indicates the format if the Localisation chain needs to give back the original
- *contentResultTarget* – value: monolingual, multilingual; indicates if the resulting translation, in the cases of several target languages, should be delivered in several monolingual content files or in a single multilingual content file

Process Data model 1

Authoring Phase

Authoring
Source QA
Enrichment
Source Terminology
Management

**Metrics/
Analysis**

Annotate w/process &
qual data
Annotate provenance of
lang. res.

Publication (CMS)

CMS reintegration
CMS Revision
Management
Publication

**Translation
Phase**

Connection to Provider

Translation

Multilingual Terminology
Management

Review

Translation QA

Post-Translation Process

General
XLIFF Connector

Pre-Translation
Human
Machine
L10n
Post-Editing

Update Linguistic
Resources
Costing and Billing

Process Data model 2

- *contentQuote* – indicates that a quoting or pricing is requested, not to perform the job
- *contentAlignment* – in case the content is to add to a Translation Memory (?)
- *contentL10N* – localize the content
- *contentI18N* – internationalize the content
- *contentDtp* – desktop publishing of content
- *contentSubtitle* – subtitling of content
- *contentVoiceOver* – voice-over of content
- *sourceRewrite: rewrite the source content (needs contentResultSource – yes)*
- *sourceReview: review the source content (needs contentResultSource – yes)*
- *sourceTranscribe: transcribe the source content (needs contentResultSource – yes)*
- *sourceTransliteration: transliterate the source content (needs contentResultSource – yes)*
- *hTranslate* – human translation
- *mTranslate* – machine translation
- *hTranscreate* – human transcreation
- *posteditQA* – human postediting of mTranslate
- *reviewQA* – human review for quality assurance only the target text, without the source text (see UNE 15038 "review"), by an expert for instance
- *reviseQA* – human revision for quality assurance examining the translation and comparing source and target (see UNE 15038 "revision")
- *proofQA* – human checking of proofs before publishing for quality assurance (see UNE 15038 "proofreading")

Scenarios for application

- C = Consumes (i.e., uses this data for its function)

- G = Generates (i.e., produced this data as part of its output)

- T = Transforms (i.e., process interprets and passes data on in a new form)

Progress-indicator

▸ ITS 2.0 must be able to convey a simple indication of the proportion of a specified process that has been completed

◦ Data model
  • progress-of-process : a process name
  • progress-indicator : 0-100
  • progress-units : (sentence|words) default: sentence

localisationCache

- Provides an indication of the status of the source and target(s) texts in a system cache for use by real-time translation, TMS, etc. to determine when retranslation is needed. A timestamp can be used to determine when the content was cached.

- **Data model**
  - **cache** – values: yes, no;
  - **scope** – values: source, target, both
  - **timestamp** – date and time (completed by Real Time Localisation System)

Discussion

- ◦ readiness

- ◦ progress indicator

- ◦ localisationCache