Simple Machine Translation, Translation Package Creation, Quality Check

ITS 2.0 in the Okapi Framework

Okapi Framework

A set of components that allow tools developers and localizers to build new localization processes or enhance existing while preserving compatibility and interoperability.

- Open-source, Cross-platform, Free

- Active community of users and developers

- See http://okapi.opentag.com/ and https://code.google.com/p/okapi/

Business Benefits of Okapi for ITS

- Provides simple and shorter development path to developers for building new ITS-enabled tools or adding ITS support to their tools.

- Provides high-level components and tools power-users can utilize to create ITS-aware processes.

- In short: it can get you ready to use ITS quickly and with little work.

ITS Parts in the Framework

- Low-level ITS Engine
- ITS-aware filters for XML, HTML5 and XLIFF
- Helper classes for ITS and XLIFF+ITS
- Microsoft Batch Translation client component
- Enrycher client component
- LanguageTool client component

See http://goo.gl/5FnfU for the detailed implementation status of the different parts.

Data Categories

- The ITS Engine implements

  ... all of them except Ruby

- The high-level components make use of

  ... most of them, for details see:

  http://www.opentag.com/okapi/wiki/index.php?title=ITS_Components

An Example

Translating XML and HTML5 documents using ITS-driven filters and a workflow that take advantage of ITS in some places.

Preparation

- Extraction
- Segmentation
- TM Leveraging
- MT Pre-translation
- Quality Check
- Translation Kit Creation

Post-processing

- Quality Check
- Translation Kit Post-processing

Business Benefits of Using ITS   (1)

- Authors can control what goes to localization from the source (Translate, Locale Filter)

- Authors can automate the feeding of helpful data for more accurate and safer localization (Domain, Localization Notes, Id Value, Preserve Space, Storage Size, Terminology, Allowed Characters, etc.)

Business Benefits of Using ITS   (2)

- Localizers have a standard way to carry the metadata throughout the project and across different tools.

- Localization can take advantage of linguistics information (Terminology, Text Analysis, MT Confidence, etc.)

- Quality control can span across the whole process, even back in the original documents (Localization Quality Issue, Localization Quality Rating)

Still To Do

- Define and handle HTML defaults

- Get MS Translator engine to preserve whitespace when Preserve Space is set

- Finalize ITS/XLIFF and XLIFF/ITS mappings

- Update of lang attributes in merged HTML

- Handle extra ITS markup when merging back

- Etc.