DITA and xml:tm

Andrzej Zydroń: azydron@xml-intl.com
DITA TC 20-December-2005
Perfect Fit

• DITA donated by IBM to OASIS
  – topic/task/concept level reuse through XML document object
• xml:tm donated by XML INTL to LISA
  – sentence level reuse through XML Namespace
• Both complement each other perfectly
• Takes the DITA reuse principle down to sentence level
• xml:tm requires a DITA environment
xml:tm

• XML based text memory
• Radically new approach to XML document localization
• First significant advance in translation memory technology
• Uses XML namespace to transparently embed contextual information
xml:tm namespace

- Text Memory namespace
- Can be mapped onto any XML document
- Vertical view of document in terms of ‘text segments’
- Can be totally transparent
xml:tm namespace

<taskbody xmlns:tm="urn:xml-Intl-tm" >
  <tm:tm>
    <context>
      <p>
        <tm:te>
          <tm:tu>
            Namespace is very flexible.
          </tm:tu>
        </tm:te>
        <tm:te>
          <tm:tu>
            It is very easy to use.
          </tm:tu>
        </tm:te>
      </p>
    </context>
  </tm:tm>
</taskbody>
xml:tm namespace

- doc
  - title
    - text
  - section
    - section
      - para
        - text
      - para
        - text
    - para
      - text
    - para
      - text
Namespace is very simple. It is easy to use.
xml:tm Text Memory

- **Author memory**
  - Maintain memory of source text
  - Authoring statistics
  - Authoring tool input

- **Translation memory**
  - Automatic alignment
  - Maintain perfect link of source and target text
  - Reduce translation costs
xml:tm DOM differencing

Source Document

```
- tu id="1"
- tu id="2"
- tu id="3"
- tu id="4"
- tu id="5"
- tu id="6"
```

Updated Source Document

```
- tu id="1"
- tu id="3"
- tu id="4"
- tu id="7"
- tu id="6"
- tu id="8"
```

- deleted
- modified
- new

origid="5"
xml:tm Author Memory

- Namespace aware DOM differencing
- Identify changes from the previous version
- Unique text unit identifiers are maintained
- Modification history
- Text units can be loaded into a database
- Authoring environment integration
Alarm

[Prolog]

[Body]

The car is equipped with a fully installed Clifford alarm system that meets category 1 Thatcham requirements. The alarm can be operated with the Spyker design key-fob. With this key fob the owner can arm/disarm the alarm and open the doors and trunk.

Figure:
Key = Arm/disarm
L = Open LH door
R = Open RH door
xml:tm Translation Memory

- The tm namespace can be used to create XLIFF files
- Automatic alignment of source and target languages
- Allows for more focused translation matching
  - Exact matching
  - Leveraged matching from document - identical text
  - Leveraged matching from database
  - Modified text unit matching
  - Non translatable text unit identification
xml:tm translation via XLIFF
xml:tm perfect alignment
xml:tm perfect matching

Updated Source Document

Perfect Matching

Matched Target Document

detected

modified

new

tu id="1"

tu id="2"

tu id="3"

tu id="4"

tu id="7"

tu id="6"

tu id="8"

requires translation

requires translation
xml:tm leveraged DB memory
xml:tm matching

Updated Source Document

Matching

Matched Target Document

tuid="1"
requires no translation

tuid="2"
requires no translation

tuid="3"
requires no translation

tuid="4"
requires no translation

tuid="7"
requires translation
requires no translation

tuid="6"
requires translation
requires no translation

tuid="8"
requires proofing

tuid="9"
requires proofing

Exact matching
non translatable
Exact matching
Exact matching
fuzzy match
doc leveraged match
DB leveraged match

non trans

new:same

DB
xml:tm benefits

• Open Standard donated by XML INTL to LISA
• Complements DITA
• Designed to work with existing standards:
  – TMX, SRX, GMX and XLIFF
• Enterprise level scalability
• Totally integrated within the XML framework
• Source text is automatically extracted and matched
• Word counts are controlled by the customer
• Text can be presented for translation via the web
• Data is merged automatically at end of translation cycle
• All memory operations are totally automated
• Can be used transparently for relay translations
• More accurate – better matching
xml:tm

• Full specification:

• Maintained by xml-intl.com
  – http://www.xml-intl.com/dtd/tm.dtd
  – http://www.xml-intl.com/dtd/tm.xsd

• Detailed article on xml:tm:

• Donated by XML INTL to Lisa OSCAR
## Web Based Translation Interface

### Oil Change

<table>
<thead>
<tr>
<th>Id</th>
<th>English, US</th>
<th>Japanese</th>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Changing the oil in your car</td>
<td>オイル交換</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>Once every 6000 kilometers or three months, change the oil in your car.</td>
<td>走行6000キロごと、または3ヶ月ごとにオイルを交換しましょう。</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>This will help maintain the engine in good condition.</td>
<td>This will help maintain the engine in good condition.</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td><strong>Source/Fuzzy Source:</strong></td>
<td><strong>Fuzzy Target:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>This will help <strong>maintain</strong> the engine in good condition.</td>
<td>エンジンを良好な状態に保ちます。</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>To change the oil:</td>
<td>オイル交換手順を示します。</td>
<td>M</td>
</tr>
</tbody>
</table>
Translation Web Services Scenario

- JSR 170 CMS
- XML + Unicode + ITS
- DITA
- TM
- TERM
- TMX
- TBX Lite
- XLIFF
- GMX/V
- GMX/C
- GMX/Q
- TBX/L
- xml:tm
- SRX
- TBX/L
Spray painting

The garage is a good place to spray paint.

Move the car out of the garage to avoid getting paint on it.

Place newspaper, cardboard, or a drop-cloth on the garage floor.

Place the object to be painted within the covered area.
The garage is a good place to spray paint.

Move the car out of the garage to avoid getting paint on it.

Place newspaper, cardboard, or a drop-cloth on the garage floor.

Place the object to be painted within the covered area.

Follow the directions on the paint can to paint the object.

Let the paint dry thoroughly before you move the object.

Remove all articles placed on the floor and dispose of safely.
Spray painting

The garage is a good place to spray paint.

Move the car out of the garage to avoid getting paint on it.

Place newspaper, cardboard, or a drop-cloth on the garage floor.

Place the object to be painted within the covered area.

Place the object to be painted on the covered area.