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
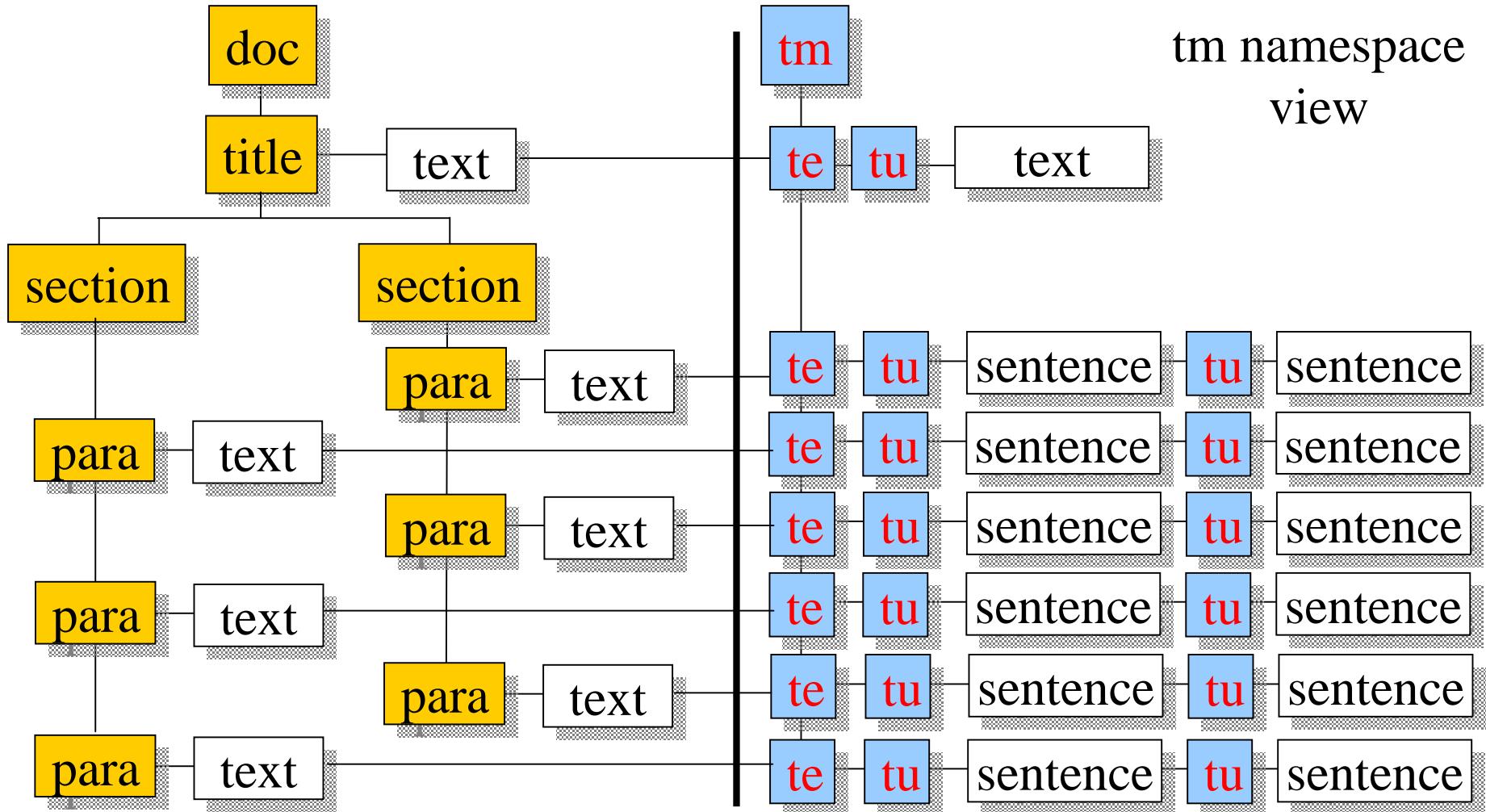
xml:tm namespace

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

xml:tm namespace

doc

title — text

section   section

section/para — text

para — text

para — text

para — text

para — text

tm

te | tu — text

te | tu | sentence | tu | sentence

te | tu | sentence | tu | sentence

te | tu | sentence | tu | sentence

te | tu | sentence | tu | sentence

te | tu | sentence | tu | sentence

te | tu | sentence | tu | sentence

tm namespace view

xml:tm namespace

original document view
```
text
```

```
<para>
Namespace is very simple. It is easy to use.
</para>
```

tm namespace view

te — tu — sentence — tu — sentence

```
<para>
  <tm:te id="e1">
      <tm:tu id="u1.1"> Namespace is very simple.    </tm:tu>
      <tm:tu id="u1.2"> It is easy to use.            </tm:tu>
  </tm:te>

</para>
```

XML-INTL
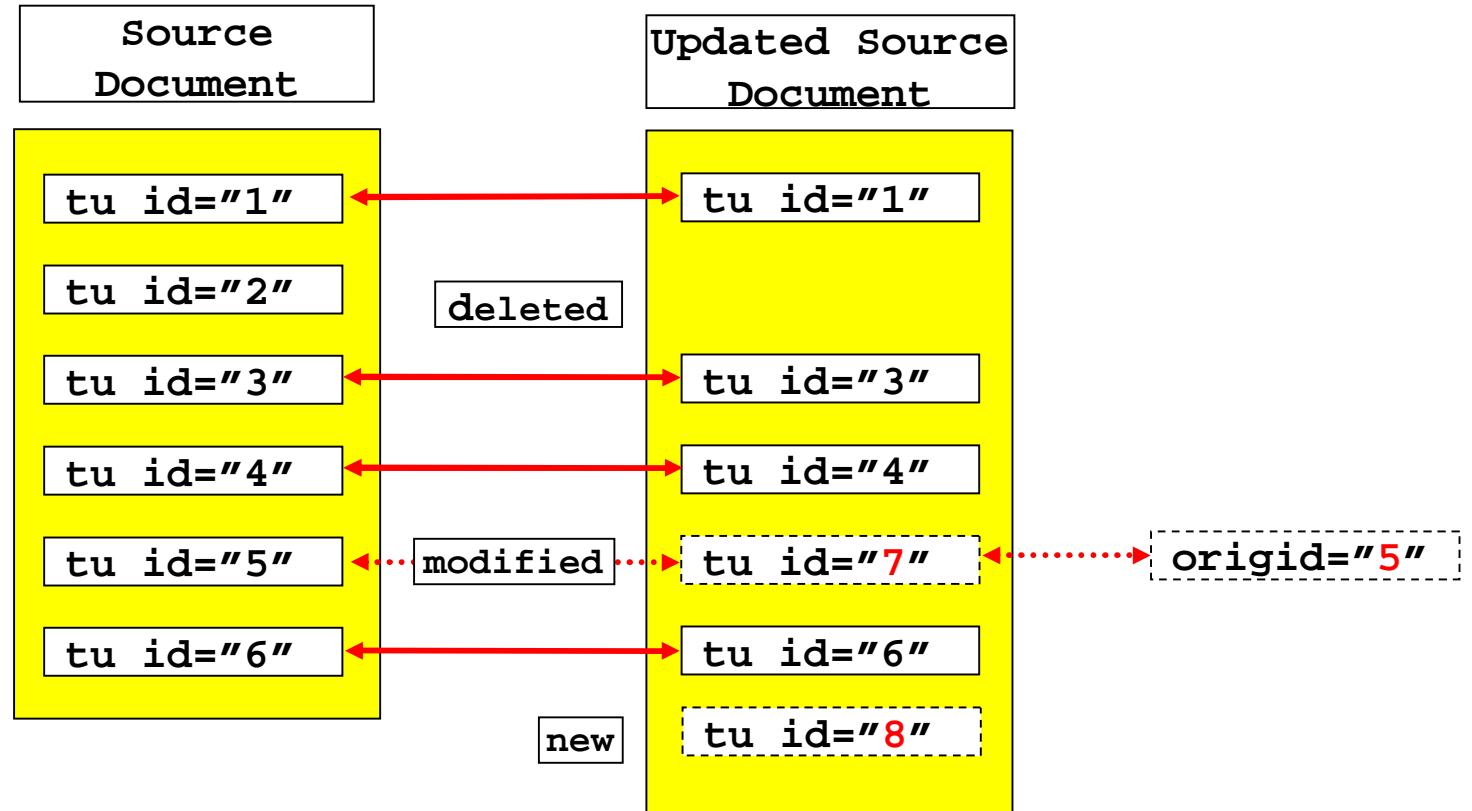
xml:tm Text Memory

- 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		Updated Source Document

tu id="1"	←——→	tu id="1"
tu id="2"	deleted	
tu id="3"	←——→	tu id="3"
tu id="4"	←——→	tu id="4"
tu id="5"	····modified····	tu id="7"	←········→ origid="5"
tu id="6"	←——→	tu id="6"
	new	tu id="8"

XML-INTL

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

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
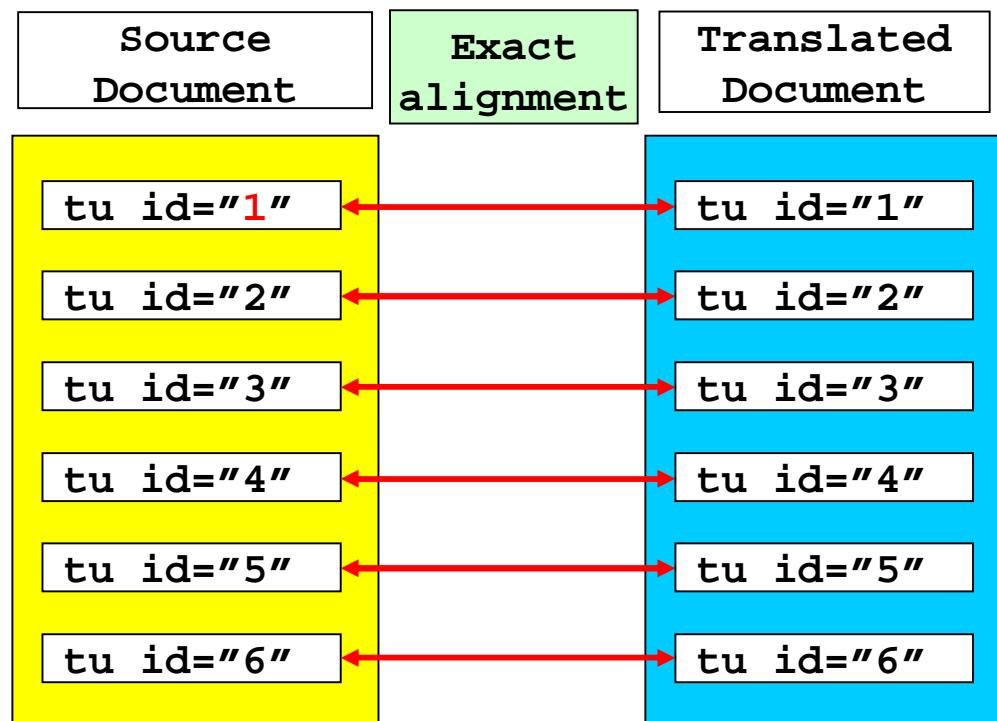
xml:tm translation via XLIFF

Source Document	XLIFF Document	Translated Document
tu id="1"	trans-unit id="1"	tu id="1"
tu id="2"	trans-unit id="2"	tu id="2"
tu id="3"	trans-unit id="3"	tu id="3"
tu id="4"	trans-unit id="4"	tu id="4"
tu id="5"	trans-unit id="5"	tu id="5"
tu id="6"	trans-unit id="6"	tu id="6"
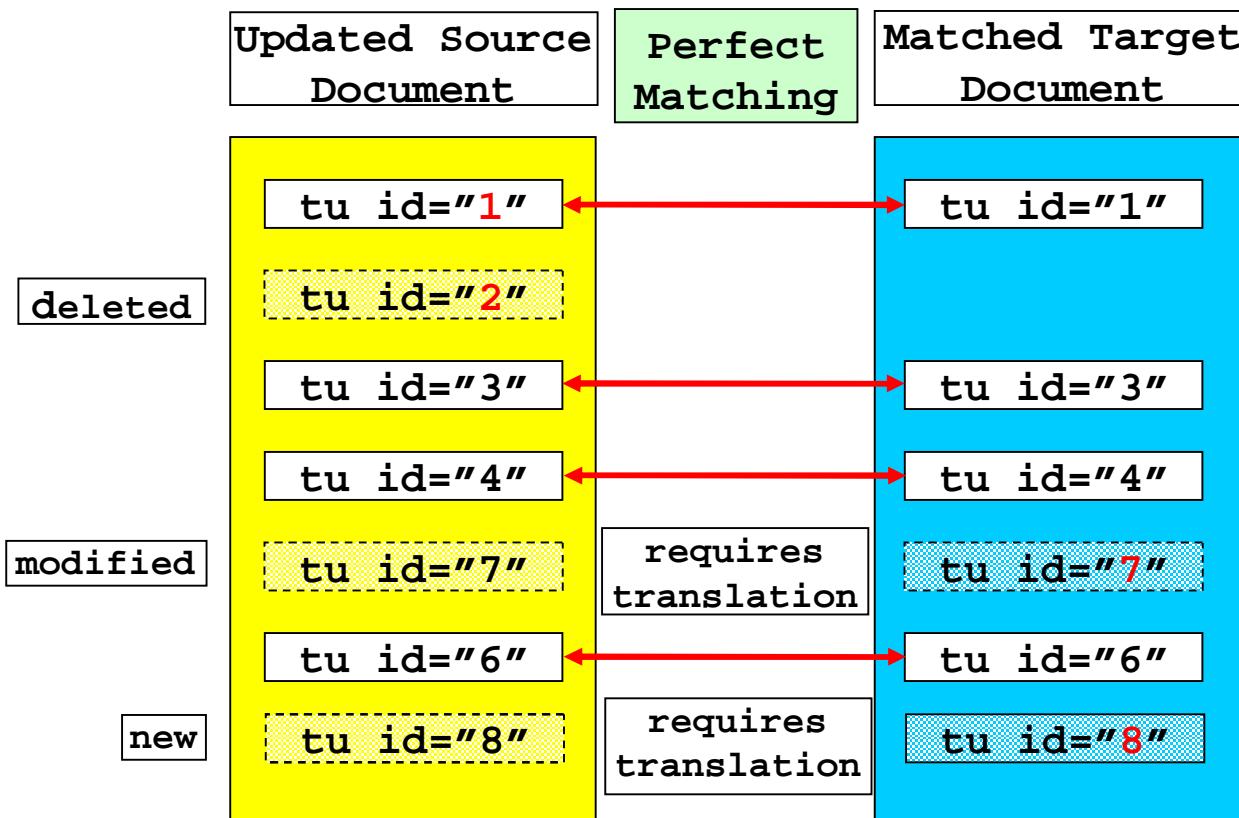
xml:tm translated document
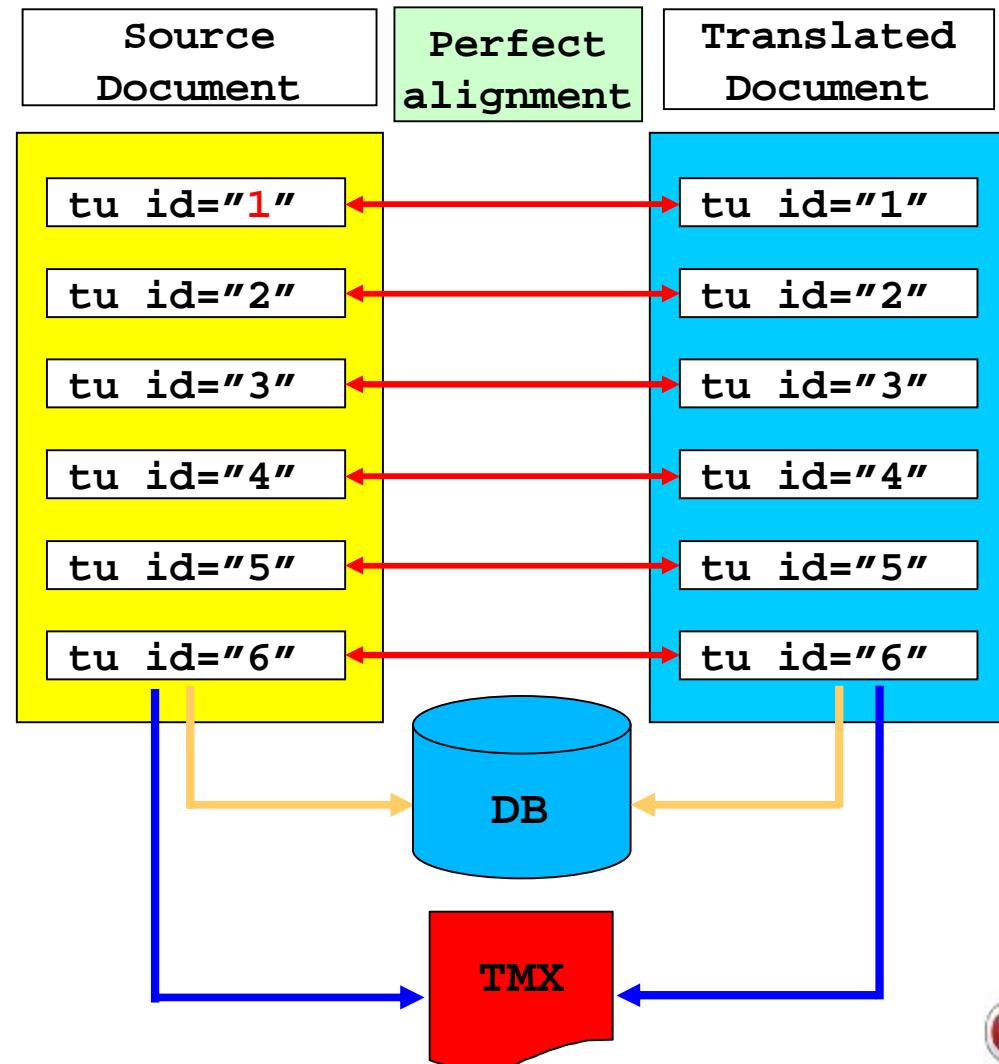
xml:tm perfect alignment

Source Document	Exact alignment	Translated Document

tu id="1"	←→	tu id="1"
tu id="2"	←→	tu id="2"
tu id="3"	←→	tu id="3"
tu id="4"	←→	tu id="4"
tu id="5"	←→	tu id="5"
tu id="6"	←→	tu id="6"

xml:tm perfect matching

Updated Source Document	Perfect Matching	Matched Target Document

tu id="1"	⟷	tu id="1"
deleted — tu id="2"		
tu id="3"	⟷	tu id="3"
tu id="4"	⟷	tu id="4"
modified — tu id="7"	requires translation	tu id="7"
tu id="6"	⟷	tu id="6"
new — tu id="8"	requires translation	tu id="8"
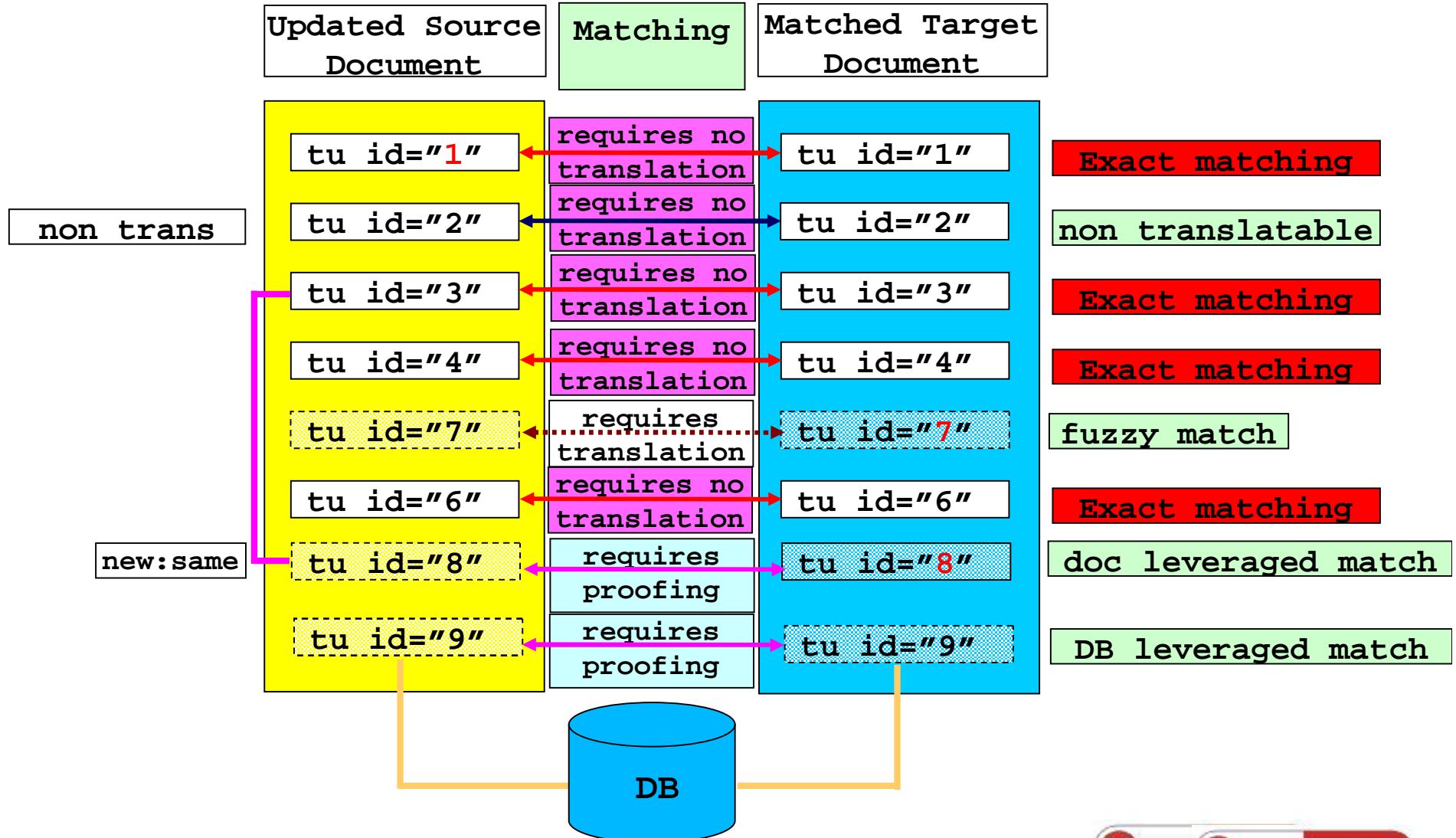
xml:tm leveraged DB memory

xml:tm matching

Updated Source Document	Matching	Matched Target Document

non trans

tu id="1"	requires no translation	tu id="1"	Exact matching
tu id="2"	requires no translation	tu id="2"	non translatable
tu id="3"	requires no translation	tu id="3"	Exact matching
tu id="4"	requires no translation	tu id="4"	Exact matching
tu id="7"	requires translation	tu id="7"	fuzzy match
tu id="6"	requires no translation	tu id="6"	Exact matching
tu id="8"	requires proofing	tu id="8"	doc leveraged match
tu id="9"	requires proofing	tu id="9"	DB leveraged match

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
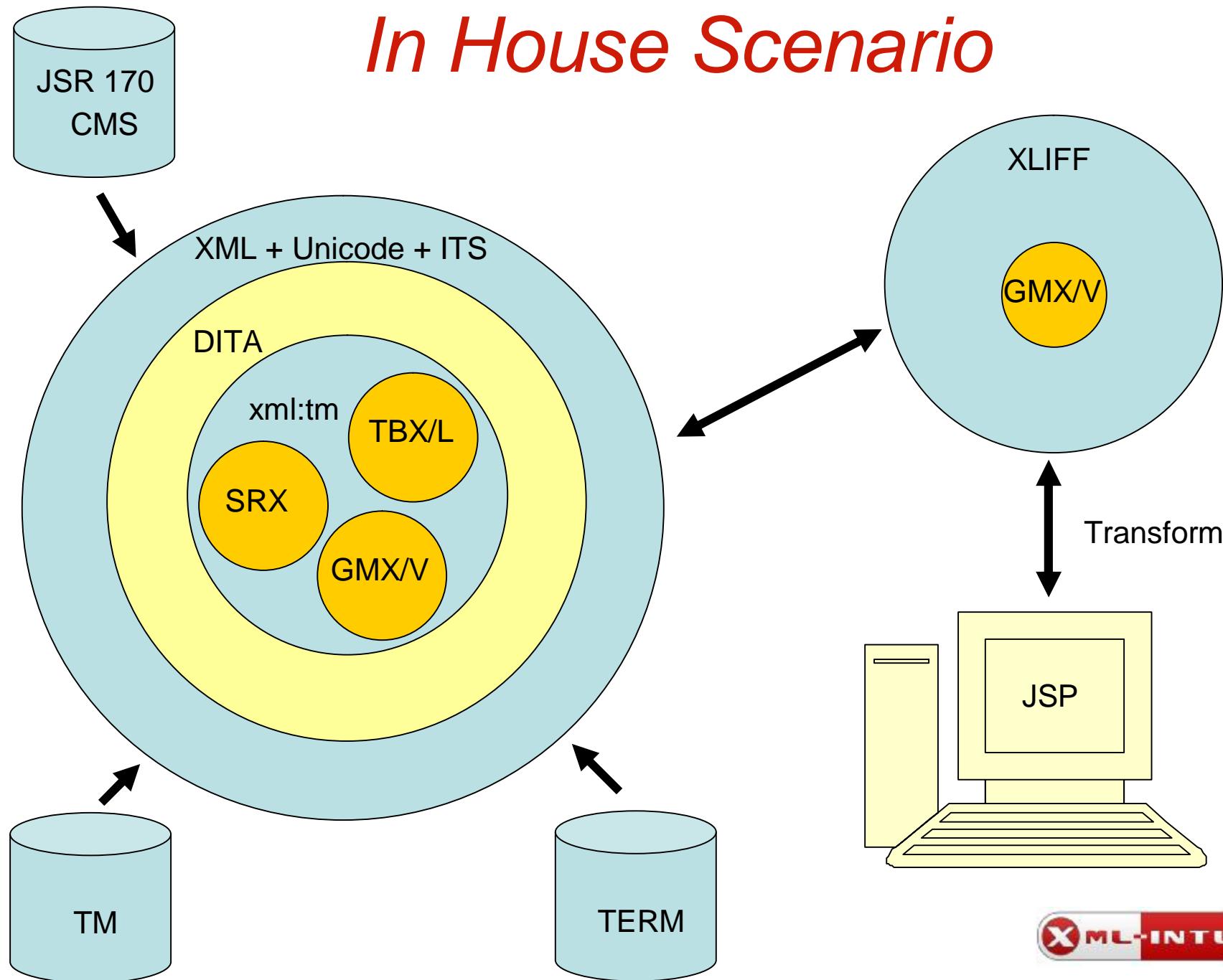
xml:tm

- Full specification:
  - http://www.xml-intl.com/docs/specification/xml-tm.html
- 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 :
  - http://www.xml.com/pub/a/2004/01/07/xmltm.html
- Donated by XML INTL to Lisa OSCAR

In House Scenario

JSR 170
CMS

XML + Unicode + ITS

DITA

xml:tm

TBX/L

SRX

GMX/V

XLIFF

GMX/V

Transform

JSP

TM

TERM

XML-INTL
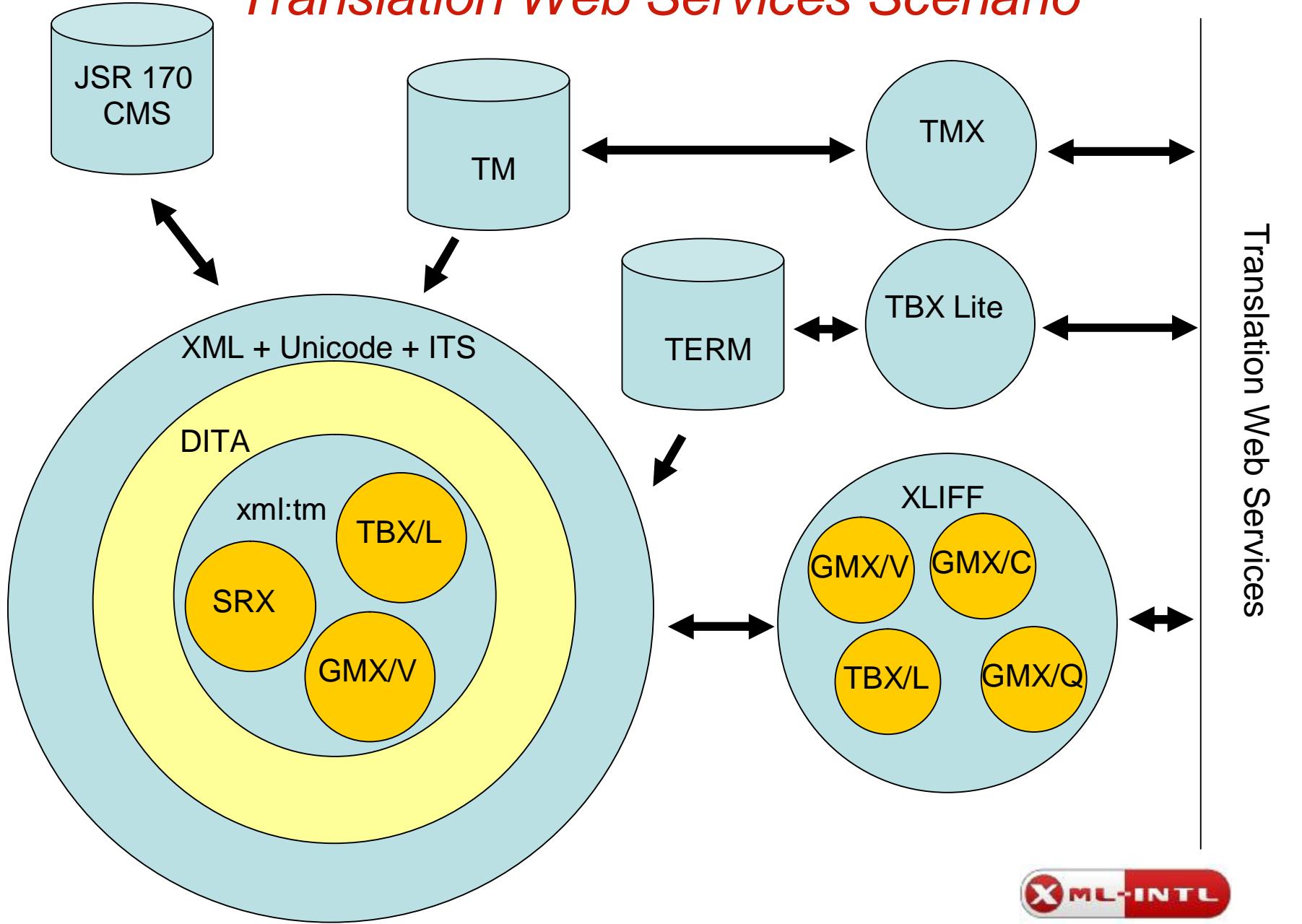
Web Based Translation Interface

Translation Web Services Scenario

JSR 170
CMS

TM

TMX

TERM

TBX Lite

XML + Unicode + ITS

DITA

xml:tm

TBX/L

SRX

GMX/V

XLIFF

GMX/V   GMX/C

TBX/L   GMX/Q

Translation Web Services

XML-INTL

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<!--
    (C) Copyright IBM Corporation 2001, 2004. All Rights Reserved.
    | This file is part of the DITA package on IBM's developerWorks site.
    | See license.txt for disclaimers and permissions.
    *
-->
<!--  ?xml-stylesheet href="../../xsl/ditaview_shell.xsl" type="text/xsl"?  -->
<task id="spraypaint" xml:lang="en-us" xmlns:tm="urn:xmlintl-tm-tags"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="../../xsd/task.xsd">
  <tm:tm date="2005-02-19T08:41:55+0000" id="d641293a-1022c5ade5d" source-language="en-US" te="9"
      version="1.0">
    <title>
      <tm:te id="e1" tu="1" version="1.0">
        <tm:tu crc="e238a5ae" id="u1.1" version="1.0">Spray painting</tm:tu>
      </tm:te>
    </title>
    <taskbody>
      <context>
        <p>
          <tm:te id="e2" tu="1" version="1.0">
            <tm:tu crc="fbde9f63" id="u2.1" version="1.0">The garage is a good place to spray
              paint.</tm:tu>
          </tm:te>
        </p>
      </context>
      <steps>
        <step>
          <cmd>
            <tm:te id="e3" tu="1" version="1.0">
              <tm:tu crc="57d66ebd" id="u3.1" version="1.0">Move the car out of the garage to avoid
                getting paint on it.</tm:tu>
            </tm:te>
          </cmd>
        </step>
        <step>
          <cmd>
            <tm:te id="e4" tu="1" version="1.0">
              <tm:tu crc="b0b48194" id="u4.1" version="1.0">Place newspaper, cardboard, or a drop-
                cloth on the garage floor.</tm:tu>
            </tm:te>
          </cmd>
        </step>
        <step>
          <cmd>
            <tm:te id="e5" tu="1" version="1.0">
              <tm:tu crc="ae8bc0a" id="u5.1" version="1.0">Place the object to be painted within the
                covered area.</tm:tu>
            </tm:te>
          </cmd>
        </step>
```

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<!--
     (C) Copyright IBM Corporation 2001, 2004. All Rights Reserved.
     | This file is part of the DITA package on IBM's developerWorks site.
     | See license.txt for disclaimers and permissions.
     *
-->
<!-- ?xml-stylesheet href="../../xsl/ditaview_shell.xsl" type="text/xsl"? -->
<task id="spraypaint" xml:lang="en-us" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:noNamespaceSchemaLocation="../../xsd/task.xsd">
 <title>Spray painting</title>
 <taskbody>
  <context>
     <p>The garage is a good place to spray paint.</p>
  </context>
  <steps>
   <step>
     <cmd>Move the car out of the garage to avoid getting paint on it.</cmd>
   </step>
   <step>
     <cmd>Place newspaper, cardboard, or a drop-cloth on the garage floor.</cmd>
   </step>
   <step>
     <cmd>Place the object to be painted within the covered area.</cmd>
   </step>
   <step>
     <cmd>Follow the directions on the paint can to paint the object.</cmd>
   </step>
   <step>
     <cmd>Let the paint dry thoroughly before you move the object.</cmd>
     Remove all articles placed on the floor and dispose of safely.
     <cmd />
   </step>
  </steps>
 </taskbody>
 <related-links>
  <link format="xml" href="../concepts/paint.xml" type="concept">
     <linktext>Paint</linktext>
  </link>
 </related-links>
 <?tm-crc 15d18371?>
 <?tm-data
 L+NZPPSyapHB0SAXTFAeQJIOv5O1kHsI3i9JxOB8TveNBkV9jhEKstAa+1k3BBW6PX9QmL5xoracFJ+H2TykqnCX+4EigA81NtlAxyGsM5p
 >
</task>
```

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE xliff (View Source for full doctype...)>
- <xliff version="1.0">
  - <file datatype="xml" original="spraypainting.diff" source-language="en-US" target-language="ja_JP" tool="xml-intl XLIFF Extract"
      xml:space="default">
    - <header>
      + <skl>
      + <count-group name="TranslationMetrics">
      + <prop-group name="ref-maps">
      </header>
    - <body>
      - <trans-unit id="t1" translate="no" ts="title" reformat="yes" xml:space="default" datatype="normal">
          <source>Spray painting</source>
          <target ts="matched">スプレー式のペンキ</target>
        </trans-unit>
      - <trans-unit id="t2" translate="no" ts="p" reformat="yes" xml:space="default" datatype="normal">
          <source>The garage is a good place to spray paint.</source>
          <target ts="matched">ガレージはスプレー式のペンキに適しています。</target>
        </trans-unit>
      - <trans-unit id="t3" translate="no" ts="cmd" reformat="yes" xml:space="default" datatype="normal">
          <source>Move the car out of the garage to avoid getting paint on it.</source>
          <target ts="matched">ペンキが付着しないように車をガレージから出します。</target>
        </trans-unit>
      - <trans-unit id="t4" translate="no" ts="cmd" reformat="yes" xml:space="default" datatype="normal">
          <source>Place newspaper, cardboard, or a drop-cloth on the garage floor.</source>
          <target ts="matched">ガレージの床に、新聞、ダンボールまたは汚れ防止の布を敷きます。</target>
        </trans-unit>
      - <trans-unit id="t5" ts="cmd" translate="yes" reformat="yes" xml:space="default" datatype="normal">
          <source>Place the object to be painted within the covered area.</source>
          <target ts="translate">Place the object to be painted within the covered area.</target>
        - <alt-trans ts="indoc-fuzzy">
            <source ts="fuzzy">Place the object to be painted on the covered area.</source>
            <target ts="fuzzy">敷き詰めた汚れ防止の上に、ペイントする物を置きます。</target>
          </alt-trans>
        </trans-unit>
```