



Integrated text and form translator for accounting report

Translation of XBRL accounting reports.

Introduction

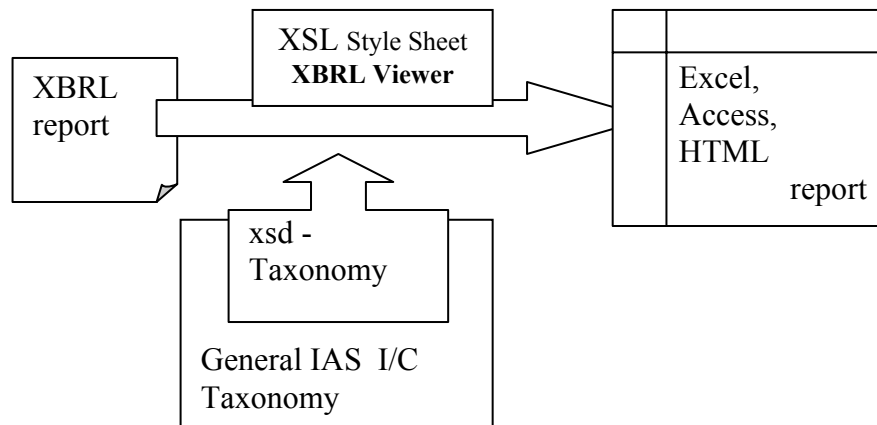
We describe in this report the integrated text and form translator for accounting reports. This stage of the work on the TRANSaccount project is based on the specialized translator (grammar and glossary) detailed in previous deliverables. The goal of this WP was to deploy the translation technologies on concrete accounting reports document. The XBRL standard (previously XFRML) has seen the 2 last years a non-expected expansion. It has been chosen as the normalized support of this application. The highly structured content of XBRL documents allows obtaining some very good description of the content and by consequent, some higher translation quality. This localization is performed using several technologies, including basic glossary substitution, translation memory or machine translation, each of them being used in a very precise context of application.

XBRL format description and localization

The XBRL format is in very fast development (<http://www.xbrl.org>), and is about to become a real exchange format for international accounting reports. Furthermore, several companies are already publishing their account reports in this format. The principle of XBRL is to represent all financial information into a special xml file, using pre-classified and normalized field (called entities). Most of these entities are shared for all accounting report. The use of an xml format allows to handle normalized validated data, and to publish them easily using some xslt operation (standard formatting operation on xml files). Generation of reference table follows this scheme: the xbrl document is converted into readable format (Excel spreadsheet, html document...) using a xsl style sheet for the formatting, and a taxonomy file describing type and labels of each entities.

Thanks to the use of this reference format, the “ontology” translation considered at the beginning of the project is not directly relevant.

Since most available material online today is in English, we have focussed the development of demonstrator on English to French language pair: the same resources can of course be used in the other way.



The different files are described in Figure 1-3.

Localizing an xbrl generated report is thus a 3-step process:

1. Localization of xbrl content
2. Localization of entities name
3. Localization of style sheet.

Step 3 allows formatting the accounting reports keeping the information organization dedicated for a special country (i.e. the usage concerning presentation of the report: position of rows, grouping of some entities...). Localization of this part is thus an experts' task. In our system we ignore this step.

Step 2, is essentially a terminology work, in figure 2, we can see that the translation of accounting information can be directly integrated to the taxonomy so that the output is produced in the give language. This task can be completed once for the reference taxonomy. In figure 2, some example of co-description of the taxonomy in both English and French is provided.

Nevertheless, it is quite frequent that each company introduces some additional entities linked with their own activity (see for instance xbrl report for Delta company). These entities will need to be localized manually in the corresponding xsd file, or translated automatically using the specialized translation engine developed in a previous stage of the TRANSaccount project.

Finally, the content of the xbrl document also needs some important localization process.



```
- <annotation>
- <appinfo>
  <xbrl:rollup to="ci:documentInformation.userDefinedDocumentInformation" weight="0" order="1" />
  <xbrl:label xml:lang="en">Edgar CIK</xbrl:label>
</appinfo>
</annotation>
</element>
- <element name="liabilitiesAndStockholdersEquity.esopPreferred" type="xbrl:monetary">
- <annotation>
- <appinfo>
  <xbrl:rollup to="ci:balanceSheet.liabilitiesAndStockholdersEquity" weight="1" order="5.5" />
  <xbrl:label xml:lang="en">Employee Stock Ownership Plan Preferred Stock</xbrl:label>
</appinfo>
</annotation>
</element>
- <element name="esopPreferred.convertiblePreferred" type="xbrl:monetary">
- <annotation>
- <appinfo>
  <xbrl:rollup to="ci:liabilitiesAndStockholdersEquity.esopPreferred" weight="1" order="1" />
  <xbrl:label xml:lang="en">Series B ESOP Convertible Preferred</xbrl:label>
</appinfo>
</annotation>
</element>
- <element name="esopPreferred.earnedCompensation" type="xbrl:monetary">
- <annotation>
- <appinfo>
  <xbrl:rollup to="ci:liabilitiesAndStockholdersEquity.esopPreferred" weight="1" order="2" />
  <xbrl:label xml:lang="en">Unearned Compensation under Employee Stock Ownership Plan</xbrl:label>
</appinfo>
</annotation>
</element>
</schema>
```

Figure 1 – Company own xsd-taxonomy

```
</appinfo>
</annotation>
</element>
- <element name="cashCashEquivalentsAndShortTermInvestments.cashAndCashEquivalents"
type="xbrl:monetary">
- <annotation>
- <appinfo>
  <xbrl:rollup to="currentAssets.cashCashEquivalentsAndShortTermInvestments"
weight="1" order="1" />
  <xbrl:label xml:lang="en">Cash and Cash Equivalents</xbrl:label>
  <xbrl:label xml:lang="fr">Trésorerie et Équivalents de Trésorerie</xbrl:label>
</appinfo>
</annotation>
</element>
- <element name="cashAndCashEquivalents.cashEquivalents" type="xbrl:monetary">
- <annotation>
- <appinfo>
  <xbrl:rollup
to="cashCashEquivalentsAndShortTermInvestments.cashAndCashEquivalents"
weight="1" order="1" />
  <xbrl:label xml:lang="en">Cash Equivalents</xbrl:label>
  <xbrl:label xml:lang="fr">Équivalents de trésorerie</xbrl:label>
```

Figure 2 - i/c xsd-taxonomy



Information & Translation Technologies

Annual Balance Sheet		Get Quarterly Data		
(values in 000's)				
Period Ending:	9/29/2001	9/30/2000	9/25/1999	9/25/1998
Current Assets				
Cash and Cash Equivalents	\$2,310,000	\$1,191,000	\$1,326,000	\$1,481,000
Short Term Investments	\$2,026,000	\$2,836,000	\$1,900,000	\$819,000
Net Receivables	\$635,000	\$1,115,000	\$824,000	\$1,137,000
Inventory	\$11,000	\$33,000	\$20,000	\$78,000
Other Current Assets	\$161,000	\$252,000	\$215,000	\$183,000
Total Current Assets	\$5,143,000	\$5,427,000	\$4,285,000	\$3,698,000
Long Term Assets				
Long Term Investments	\$128,000	\$786,000	\$0	\$0
Fixed Assets	\$564,000	\$313,000	\$318,000	\$348,000
Other Assets	\$186,000	\$277,000	\$558,000	\$243,000
Total Assets	\$6,021,000	\$6,803,000	\$5,161,000	\$4,289,000
Current Liabilities				
Accounts Payable	\$1,518,000	\$1,933,000	\$1,549,000	\$1,520,000
Total Current Liabilities	\$1,518,000	\$1,933,000	\$1,549,000	\$1,520,000
Long Term Debt	\$317,000	\$300,000	\$300,000	\$954,000
Deferred Liability Charges	\$266,000	\$463,000	\$208,000	\$173,000
Total Liabilities	\$2,101,000	\$2,696,000	\$2,057,000	\$2,647,000
Stock Holders Equity				
Preferred Stocks	\$0	\$76,000	\$150,000	\$150,000
Common Stocks	\$1,693,000	\$1,502,000	\$1,349,000	\$633,000
Retained Earnings	\$2,260,000	\$2,285,000	\$1,499,000	\$898,000
Other Equity	(\$33,000)	\$244,000	\$106,000	(\$39,000)
Total Equity	\$3,920,000	\$4,107,000	\$3,104,000	\$1,642,000

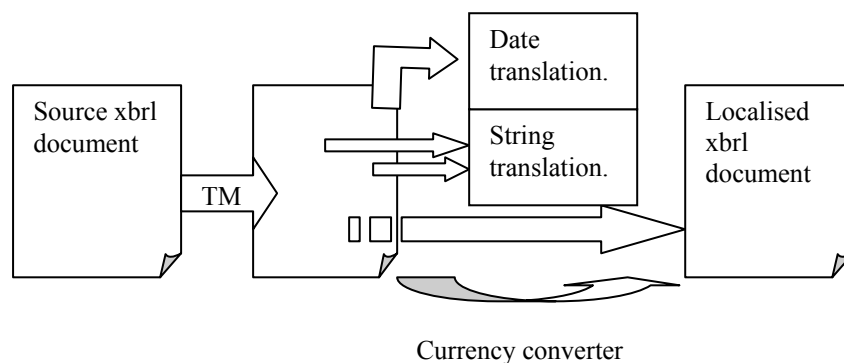
Figure 3 - Conversion of xbrl report to Excel spreadsheet



Information & Translation Technologies

Localization of XBRL report:

Localization of xbrl report has required analysing and translating sub-items of the xml file (we remind that xbrl is a special xml file) according to the following figure:



The xml parsing distinguishes the following XML elements and attributes for localization:

- Date and Period areas: using a special module, “translations” of date and period attributes are converted to the closer format suitable for the target language. This module is based on a large coverage source description of these fields. These areas are identified in the xsd file in the attribute `period="xx"`
- Currency: depending on user configuration and source/target language, some currency translation may be applied; the process can at the same time perform some currency conversion. This currency conversion information is added to the source information, and do not substitute this information `$332000 → $332000 (xx Fr)`
- Label Localization: the translated taxonomy is the first source of this localization. Translated taxonomy is integrated in the xsd file. At translation time, the current xsd are analysed and all non-translation entity labels are translated using customized translation engine.
- Proper name, and/or address: since information in the xbrl document is very structured, it is also possible to extract and handle text fields not requiring translation or rearrangement such as proper names. This is performed by manual identification of DNT (do not translate) entities.
- Free text translation: this part translates all included free text in reports (comments, introduction, notes, general presentation).



Implementation

Implementation of the preceding mechanism is based on the use of an adapted xml parser in the following steps:

- Build the internal tree corresponding to xbrl document
- Label Localization: Identify in the tree of already localized items by xsd parsing, machine translation on other fields.
- Extract of date/currency and apply appropriate mechanism
- Translation of free part using main engine with specialized resources
- Dump of internal tree to regenerate localized xbrl document.

Remark that xbrl document are intrinsically multilingual, using the `xml:lang="en"` attribute. The localized document is still containing source information, as well as localized information.

XBRL sample

```
type="ci:statements" entity="Delta" units="ISO4217:USD" scaleFactor="6" precision="10" decimalPattern="#.##"
formatName=""
- <group type="ci:statements.balanceSheet">
- <group type="ci:cashCashEquivalentsAndShortTermInvestments.cashAndCashEquivalents">
  <label href="xpointer(..)" xml:lang="en">Cash and cash equivalents</label>
  <item period="2000-06-30">1252</item>
  <item period="1999-06-30">1124</item>
</group>
- <group type="ci:cashCashEquivalentsAndShortTermInvestments.ShortTermInvestments">
  <label href="xpointer(..)" xml:lang="en">Short-term investments</label>
  <item period="2000-06-30">493</item>
  <item period="1999-06-30">19</item>
</group>
- <group type="ci:currentAssets.receivablesNet">
  <label href="xpointer(..)" xml:lang="en">Accounts receivable, net of allowance for uncollectible accounts of
  $34 at June 30, 2000 and $30 at June 30, 1999</label>
  <item period="2000-06-30">739</item>
  <item period="1999-06-30">662</item>
</group>
- <group type="ci:currentAssets.deferredIncomeTaxesCurrentPortion">
  <label href="xpointer(..)" xml:lang="en">Deferred income taxes</label>
  <item period="2000-06-30">356</item>
  <item period="1999-06-30">403</item>
</group>
- <group type="ci:currentAssets.prepaidExpenses">
  <label href="xpointer(..)" xml:lang="en">Prepaid expenses and other</label>
  <item period="2000-06-30">506</item>
  <item period="1999-06-30">524</item>
</group>
- <group type="ci:assets.currentAssets">
  <label href="xpointer(..)" xml:lang="en">Total current assets</label>
  <item period="2000-06-30">3346</item>
  <item period="1999-06-30">2672</item>
</group>
```

In this sample, we show in red plain circle labels translated using direct multilingual glossary, in dashed green circle, a free text area. The `<item period>` are simple dates (in this case, the translation to French, it will only invert order of day/month/year). And in the item element, the currency, here dollar, will be or not translated to some other currency depending on user choice.