Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge across Europe

Presented for the Monnet Consortium by Thierry Declerck, DFKI GmbH
• **Industrial Organizations**
  - BeInformed, The Netherlands ([http://www.beinformed.nl/](http://www.beinformed.nl/))
  - XBRL-EUROPE, Belgium ([http://www.xbrl.org/eu/](http://www.xbrl.org/eu/))

• **Research Institutes**
  - DERI, Ireland ([http://www.deri.ie/](http://www.deri.ie/)), coordinator
  - DFKI, Germany ([www.dfki.de](http://www.dfki.de))
  - UPM, Spain ([http://www.upm.es/institucional](http://www.upm.es/institucional))
  - UniBi, Germany ([http://www.uni-bielefeld.de](http://www.uni-bielefeld.de))
Main Objectives

- **Multilingual Ontologies for Networked Knowledge**
  - Linguistically enriched knowledge representation
  - Multilingual access to structured/networked knowledge: ontologies, knowledge bases, linked data
  => enabling the *Multilingual Web*

- See the Multilingual Web project,
  http://www.multilingualweb.eu/:
  „The MultilingualWeb project is exploring standards and best practices that support the creation, localization and use of multilingual web-based information.“
• As we will see, the Monnet project investigates a specific solution to this issue. We first give examples of web pages that convey multilingual information, but that are not at the end instances of the Multilingual Web, in our sense.
 Budapesti Értéktőzsde
Both the Hungarian and the English pages are parallelized and give good localized information (including on the values of certain descriptors), which can be extracted and transformed in machine-processable code (see next slide).

But there is no possibility to address semantic queries to those pages or to consult those if the user does not understand one of both languages.

- This calls for the applications of semantic technologies that extract information and store these in language-independant way, as a base for generation in another language.
Using English as a pivot language?

• Looking at similar pages in German, Italian and Spanish, one can observe that English is used in all the cases as the second language, and that similar bi-lingual terminology and information can be extracted.

• But they make use of many variations in the terms used, so that we can not use English as the intermediate language for localizing from Hungarian to, for example, Spanish.

• The Monnet approach is proposing a combination of Semantic Web and Translation technologies for solving this problem.
Toward a multilingual Semantic Web

• Monnet is handling information at a semantic level
  – Abstracting from language and form
  – Cross-lingual Ontology-based Information Extraction and Knowledge
    • integration,
    • aggregation,
    • querying,
    • presentation

• The Semantic Web as such is not proposing ways to deal with natural language and multilingualism, and a first step in Monnet consists in proposing a model for the representation of natural language expressions used in ontologies in the Semantic Web
Make use of linguistically described multilingual terms in Ontologies

The OTR Model, as described in [2]

ISO Linguistic Annotation Framework (including data categories)
Morpho-Syntax Description
Gender: M, F, N, ...
Number: Sg, Pl, ...
Lemma (pointing to lexicon)
POS: N, V, P, A, ...
Syntax
Constituency
NP, VP, PP, ...
Dependency
Head, Modifier, Specifier, ...

Domain Ontology
(Financial Reporting)
Class1
Class1_1 (Is_A Class1)
Class1_2 (Is_A Class1)
...
Class2
Class2_1 (Part_Of Class2)
Class2_2 (Part_Of Class2)
...

Tokenized Terms Classes (in four languages)
Term1 (NL): [t, Grundlagen] [t, voor] [t, financiën] [t, verlaggeving]
Term1 (EN): [t, Accounting] [t, Policies]
Term1 (DE): [t, Bilanzierungs-] [t, und] [t, Bewertungsmethoden]
Term1 (ES): [t, Politicas] [t, Contables]
Term2 (NL): Opname van actuariële winsten en verliezen in het kader van toegerekend-pensioenregelingen, Grondslag
Term2 (EN): [t, Accounting] [t, Policy] [t, for] [t, Recognition] [t, of] [t, Account]
[t, Gains] [t, and] [t, Losses] [t, Defined] [t, Benefit] [t, Plan]
Term2 (DE): Bilanzierungs- und Bewertungsmethoden zur Erfassung
Versicherungsmathematischer Gewinne und Verluste aus leistungsorientierten Plänen
Term2 (ES): Política Contable para el Reconocimiento
de Garanías y Pérdidas Actuariales Planes de Beneficios Definidos
Monnet Architecture and Components

- Corpus Service
- Lexicalization Service
- Information Extraction Service
- Knowledge Access and Presentation Service
- Ontology
- Localization Service

Translator and Expert Interactions:
- $100
The lemon model, for encoding lexicalized ontologies
Term Analysis for Machine Translation

minimum finance lease payments receivable …

Google-Translate: minimale Finanzierung uneinbringliche

IFRS: im Rahmen von Finanzierungs-Leasingverhältnissen zu erhaltende Mindestleasingzahlungen

term analysis – using: IFRS, SAPterm, IATE, DBpedia, ...

minimum finance lease payments receivable

sub-term translation – using: IATE, DBpedia, Leo, ...

mindest Finanzierungsleasing Zahlungen erträge

term synthesis – using: rules? statistical models?

Mindestfinanzierungsleasingzahlungserträge

Google-Translate: minimum finance lease payment income
Scheduled maturities of minimum lease payments outstanding at December 31, 2007, expressed as a percentage of the total, are approximately: 2008, 50 percent; 2009, 30 percent; 2010, 15 percent; 2011, 5 percent; and 2012 and beyond, 1 percent.

Barclay’s Annual Report 2009
The allowance for uncollectable finance lease receivables included in the allowance for impairment amounted to £321m at 31st December 2009 (2008: £189m).

The total of future minimum sublease payments to be received under non-cancellable subleases at the balance sheet date is £147m (2008: £158m).
Outcomes

- An *Ontology Lexicalization Service* to extract lexical expressions used in labels of ontologies, and to create an ontology lexicon automatically enriched with linguistic information.
- An *Ontology Localization Tool* to create an ontology lexicon in a target language from an ontology lexicon in a source language, semi-automatically.
- A *Cross-Lingual Ontology-based Information Extraction System* (CLOBIE) to leverage multilingual ontology lexicons to extract information from text and populate ontologies.
- A *Cross-Lingual Query & Presentation System*, which uses multilingual ontology lexicons to enable quick customization of knowledge access systems to many natural languages. The knowledge being here stored in populated ontologies.
Use Cases

- The project is validating its approach to enabling the multilingual web in the context of two use cases:
  - in the field of e-Government
  - in the field of financial and business information
- Focus in this talk is on 2., called IFRS-XBRL use case, which aims at enabling the search and the report creation of financial information and business service descriptions in the language of choice of the users.
IFRS-XBRL Use case

• Objective: Enable search and report creation in language of choice of financial information
  – Business intelligence on European companies, involving semantic-level analysis of business reports

• Impact: An integrated solution to providing semantic-level access to financial information across languages
  – The prototype will allow a financial analyst to search for information by filling in structured search forms localized to his/her own language. The results will be presented in terms of charts, diagrams, results lists etc. localized to the preferred language of the user
Data Sources

- **Structured sources:**
  - Publicly available balance sheets in structured format (for example in German at http://www.bundesanzeiger.de/);
  - Short company profiles (e.g., from Business Registers, Stock Exchange web pages, etc.);
  - Wikipedia Infoboxes; and
  - XBRL instance documents (the Belgian National Bank has published online all the XBRL reports of Belgian companies).

- **Semi-structured:**
  - Longer company profiles;
  - Imprint information on company web pages;
  - Running tickers on company information.

- **Unstructured:**
  - Annexes to balance sheets in annual reports of companies;
  - Newspapers;
  - Specialized web pages etc.
Main beneficiaries:

1. The translators of the XBRL taxonomies (who will profit from the effort reduction yielded by using our automatic ontology localization functionality)
2. The users of XBRL taxonomies (e.g. financial analysts), who will be able to see the information formalized by XBRL instance documents or extracted from unstructured resources in accordance to the XBRL taxonomies in their preferred language.
XBRL and IFRS

- XBRL (eXtensible Business Reporting Language) is an XML-based mark-up language for the exchange of business information, including financial reporting. Its use is being nowadays mandated by a growing number of regulatory bodies and stock exchanges around the world.

- IFRS (International Financial Reporting Standards) is a standard developed by the International Accounting Standards Board (IASB). IFRS is also available as a taxonomy encoded in XBRL.
### Available IFRS Modules

<table>
<thead>
<tr>
<th>Module</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRS 1</td>
<td>First-time Adoption of International Financial Reporting Standards</td>
</tr>
<tr>
<td>FRS 2</td>
<td>Share-based Payment</td>
</tr>
<tr>
<td>FRS 3</td>
<td>Business Combinations</td>
</tr>
<tr>
<td>FRS 4</td>
<td>Insurance Contracts</td>
</tr>
<tr>
<td>FRS 5</td>
<td>Non-current Assets Held for Sale and Discontinued Operations</td>
</tr>
<tr>
<td>FRS 6</td>
<td>Exploration for and Evaluation of Mineral Resources</td>
</tr>
<tr>
<td>FRS 7</td>
<td>Financial Instruments: Disclosures</td>
</tr>
<tr>
<td>FRS 8</td>
<td>Operating Segments</td>
</tr>
<tr>
<td>IAS 1</td>
<td>Presentation of Financial Statements</td>
</tr>
<tr>
<td>IAS 2</td>
<td>Inventories</td>
</tr>
<tr>
<td>IAS 7</td>
<td>Statement of Cash Flows</td>
</tr>
<tr>
<td>IAS 8</td>
<td>Accounting Policies, Changes in Accounting Estimates and Errors</td>
</tr>
<tr>
<td>IAS 10</td>
<td>Events after the Reporting Period</td>
</tr>
<tr>
<td>IAS 11</td>
<td>Construction Contracts</td>
</tr>
<tr>
<td>IAS 12</td>
<td>Income Taxes</td>
</tr>
<tr>
<td>IAS 15</td>
<td>Property, Plant and Equipment</td>
</tr>
<tr>
<td>IAS 17</td>
<td>Leases</td>
</tr>
<tr>
<td>IAS 18</td>
<td>Revenue</td>
</tr>
<tr>
<td>IAS 19</td>
<td>Employee Benefits</td>
</tr>
<tr>
<td>IAS 20</td>
<td>Accounting for Government Grants and Disclosure of Government Assistance</td>
</tr>
<tr>
<td>IAS 21</td>
<td>The Effects of Changes in Foreign Exchange Rates</td>
</tr>
<tr>
<td>IAS 23</td>
<td>Borrowing Costs (2007-03-01)</td>
</tr>
<tr>
<td>IAS 24</td>
<td>Related Party Disclosures</td>
</tr>
<tr>
<td>IAS 25</td>
<td>Statement of changes in net assets available for benefits</td>
</tr>
<tr>
<td>IAS 27</td>
<td>Consolidated and Separate Financial Statements</td>
</tr>
<tr>
<td>IAS 28</td>
<td>Investments in Associates</td>
</tr>
<tr>
<td>IAS 29</td>
<td>Financial Reporting in Hyperinflationary Economies</td>
</tr>
<tr>
<td>IAS 31</td>
<td>Interests in Joint Ventures</td>
</tr>
<tr>
<td>IFRS 7</td>
<td>Financial Instruments: Disclosures</td>
</tr>
<tr>
<td>IFRS 8</td>
<td>Operating Segments</td>
</tr>
</tbody>
</table>

[IFRS Modules](http://www.xbrl-ifrs.org/ITMM/)
A Financial Reports: duration vs. Instant

FINANCIAL STATEMENTS IFRS

Consolidated Income Statements of SAP Group
for the years ended December 31,

<table>
<thead>
<tr>
<th>€ millions, unless otherwise stated</th>
<th>Note</th>
<th>2009</th>
<th>2008</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software revenue</td>
<td></td>
<td>2,807</td>
<td>3,608</td>
<td>3,407</td>
</tr>
<tr>
<td>Support revenue</td>
<td></td>
<td>5,285</td>
<td>4,602</td>
<td>3,852</td>
</tr>
<tr>
<td>Subscription and other software-related service revenue</td>
<td></td>
<td>308</td>
<td>258</td>
<td>182</td>
</tr>
<tr>
<td>Software and software-related service revenue</td>
<td></td>
<td>8,198</td>
<td>8,466</td>
<td>7,441</td>
</tr>
<tr>
<td>Consulting revenue</td>
<td></td>
<td>2,074</td>
<td>2,498</td>
<td>2,221</td>
</tr>
<tr>
<td>Training revenue</td>
<td></td>
<td>273</td>
<td>434</td>
<td>410</td>
</tr>
<tr>
<td>Other service revenue</td>
<td></td>
<td>85</td>
<td>107</td>
<td>113</td>
</tr>
<tr>
<td>Professional services and other service revenue</td>
<td></td>
<td>2,132</td>
<td>2,039</td>
<td>2,744</td>
</tr>
<tr>
<td>Other revenue</td>
<td></td>
<td>42</td>
<td>70</td>
<td>71</td>
</tr>
<tr>
<td>Total revenue</td>
<td></td>
<td>(5)</td>
<td>10,672</td>
<td>11,575</td>
</tr>
</tbody>
</table>

Consolidated Statements of Financial Position of SAP Group
as at December 31,

<table>
<thead>
<tr>
<th>€ millions</th>
<th>Note</th>
<th>2009</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash and cash equivalents</td>
<td></td>
<td>1,884</td>
<td>1,280</td>
</tr>
<tr>
<td>Other financial assets</td>
<td>(13)</td>
<td>406</td>
<td>566</td>
</tr>
<tr>
<td>Trade and other receivables</td>
<td>(14)</td>
<td>2,546</td>
<td>3,178</td>
</tr>
<tr>
<td>Other non-financial assets</td>
<td>(15)</td>
<td>147</td>
<td>176</td>
</tr>
<tr>
<td>Tax assets</td>
<td>(11)</td>
<td>192</td>
<td>399</td>
</tr>
<tr>
<td>Total current assets</td>
<td></td>
<td>5,265</td>
<td>6,571</td>
</tr>
</tbody>
</table>

Such Instances of financial reports exists at most in 2 languages!
High Level Model of XBRL by Charles Hoffman is licensed under a Creative Commons Attribution 3.0 United States License.
<element name="Assets" id="ifrs_Assets" periodType="instant" balance="debit" abstract="false" substitutionGroup="item" type="monetaryItemType"/>

Labels
Cash and Equivalents
現金及び現金等

References
FAS 12, Section A, Paragraph 1(b) III

Formulas (rules)
IF Cash and Equivalents < 0 THEN Classify as Liability

Presentations

Calculations
Cash and Equivalents = Cash + Petty Cash + Bank account

Definitions
Is an "Asset"; Audit Risk is "High"
Searching through the IFRS Taxonomy in XBRL
Translation of the Core IFRS Standards, as described by IFRS

- **Key Term Extraction**: IASC Foundation extracts key terms from the IFRS standards.
- **Key Term Translation**: The key terms are translated by the translator.
- **Agreement On Translation**: Translated key terms are agreed upon by the Committee.
- **IFRS Draft Translation**: Translator uses key terms and existing IFRS reference material to translate IFRSs.
- **Review & Finalization**: Committee reviews draft translation for accuracy and consistency, and text is finalised.

Translated into about 30 languages

Termbase makes use of
### Comparing Translation of Text to Translation of XBRL Taxonomy

<table>
<thead>
<tr>
<th>Model in Workbench</th>
<th>Metadata about elements</th>
<th>Context seen by translator</th>
<th>Context used for suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Sequence of text/segments</td>
<td>Text type: paragraph, header, caption, other?</td>
<td>Entire text</td>
</tr>
<tr>
<td>XBRL Taxonomy</td>
<td>Graphs of labels (sequentialized somehow)</td>
<td>Type of label; Concept attributes</td>
<td>Locations in graphs; reference to standard</td>
</tr>
</tbody>
</table>

**Hypothesis:** Given the taxonomy, English labels, and the translation of the IFRS standards into a language, the XBRL labels for the language can be automatically deduced with a very high accuracy. The translator would, however, need to check them. The IFRS termbase might aid this task.
Example IFRS Taxonomy Label and its Translation

- Minimum finance lease payments receivable, at present value, end of period later than one year and not later than five years
- Im Rahmen von Finanzierungs-Leasingverhältnissen zu erhaltende Mindestleasingzahlungen, zum Barwert, länger als ein Jahr und bis zu fünf Jahren bis zum Ende der Periode

Example of inconsistency in IFRS 2009 Taxonomy Labels:

- ‘end of period later than one year and not later than five years’
  - occurs as a segment in 8 labels.
- It is translated into German 6 times as:
  - ‘länger als ein Jahr und bis zu fünf Jahren bis zum Ende der Periode’
- 2 times as:
  - ‘mehr als ein Jahr und bis zu fünf Jahren bis zum Ende der Periode’
Information Extraction (IE) can find values in text and relate them to concepts, so that the values can be checked.

Similarly, IE can help with detailed tagging.

Example text about derivatives

- The fair value hedge has a notional amount of $250 million, and hedges approximately 86% of the $292 million of outstanding senior notes maturing in September 2011.

IE equates the three underlined values with XBRL tags to create three different facts:
- NotionalAmountOfInterestRateFairValueHedgeDerivatives has the value $250 million.
- PercentageOfDebtHedgedByInterestRateDerivatives has the value 86%.
- SeniorNotes has the value $292 million.

The facts might also have the context September 2011

Information about the taxonomy location (etc.) of the concepts may help confirm the interpretation.
Monnet involved in xEBR
European Business Registers
Mapping between taxonomies

Registers - Information Providers
(XEU BR WG Members)

- National Bank of Belgium (Belgium)
- Eogs / DCCA (Denmark)
- Registrite ja infosüsteemide Keskus eRik (Estonia)
- Bilans Service - Infogreffe (France)
- Bundesanzeiger (Germany)
- Infocamere (Italy)
- RSCL (Luxembourg)
- Kamer van Koophandel (Netherlands)
- Informa DB – Colegio de Registradores (Spain)
- Bolagsverket (Sweden)
- Companies House (United Kingdom)
- EBR (Europe)
- GBR (Global)
- IASCF
- Bank of Spain
- Software – Audit – Consulting
An example of Mapping

```
Taxonomy 1       XEUBR Core Ref Taxo       Taxonomy 2

Address ← Address ← Address

Line 1 ← Line 2 ← Place

Line 3 ← ZIP ← Street

ZIP ← City ← Number

City ← City ← Complement

ZIP ← City
```
xEBR European Business Registers
Mapping between taxonomies

Core Concepts

- Dutch GAAP
- Belgian GAAP
- German GAAP
- Spanish GAAP
- French GAAP
- UK GAAP
Summary

• There are various ways to ensure interoperability of information in the web
  – By mapping knowledge representation systems (not at the core of MONNET)
  – My making knowledge representation systems multilingual, in order to support real interaction with the European citizen

• Thanks to the Montific project and for you attention
  – Contact: declerck@dfki.de