Minutes of ECIMF meeting, 16/10/2001

Meeting started: 09:40
Meeting ended: 17:15

Attendees
Andrzej Bialecki (Chair)
Andrea Boggiano **
Martin Bryan
Prem Couture *
Paul Metusala Dikobe
Kevin Doyle **
Mike Lambert **
Heiner Stuckenschmidt **
Tony Fletcher (REA presentation)
* Not morning session
** Not afternoon session

Apologies for absence
The rail strikes made it impossible for some to attend. Among those unable to attend were:
Mr. Frederic Camuzard from Motorola
Prof. Asuncion Gomez-Perez, University of Madrid

Martin Bryan was appointed as secretary for the meeting.

As there were no comments on the minutes of the last meeting these were accepted.

The outstanding items were reviewed. It has proved difficult to obtain feedback on the project plan from the OAG and OASIS, but comments have been received from RosettaNet. Contact has been made with Prof. Bill McCarthy, one of the developers of REA, who can help to ensure that the ECIMF specifications remain aligned with REA.

RosettaNet is working on techniques for transitioning from their PIPs to ebXML. Mr Coleman the Chief Architect for RosettaNet was unable to attend this meeting but it is hoped that he will be able to provide feedback on the guidelines.

Work on the semantic translation mechanism for ECIMF has been slow because most of the currently available techniques are very difficult for business SMEs to understand. An easily understandable model for semantic translation is needed. Work on the mapping methodologies has been easier, and most of the current work has been in this area. The addition of the business context layer has necessitated changes in the other sections, but these have yet to me made. Andrzej needs help with editing the text.

The project has a website at www.ecimf.org, which includes a discussion group, though this is currently not being used.

RosettaNet is hoping to provide a scenario for the Proof-of-Concept document based on the updating of PIPs to ebXML. Andrzej is to convert the Powerpoint slides on the work he has been doing into a mapping scenario for inclusion into the POC document. Work using Conzilla for translation has proved unsuccessful to date. (More was said on this later.)

Presentations

Business Context Equivalence – Andrzej Bialecki, WebGiro

The subtitle of this presentation was Using REA and UMM for Interoperability. ECIMF is based on top-down analysis, with an iterative process based on Business Process Mediation, Semantic Translation and Syntax Mapping. However, it needs to start from business goals. The first draft of the Framework Integration Guide (FIG) lacked a properly defined model for that area. This led to a new layer, Business Context, being added to the ECIMF model.

The Resource Event Agent (REA) ontology is used to as the underlying meta-model for a number of methodologies, including UMM and ebXML. REA is expressed in business terms, such as Economic Resource, Economic Commitment, Economic Event, and Economic Agent. There is a Knowledge Infrastructure and an Operational Infrastructure. Economic Events occur in pairs. Two events are needed to define a reciprocal commitment known as an agreement. Events that take place at the same time and place are known as congruent events. Alternatively events can take place at separate times (e.g. goods despatch and goods reception).

Recent extensions to REA have defined a REA Enterprise Script, which allows for dynamic processes as well as static ones. Here Business Processes such as stock flows and revenue processes, are linked by exchanges, which consist of a Give and Take cycle. The set of ordered tasks that define the exchange are referred to as a Recipe.

The UN/CEFACT Unified Meta-Model (UMM) Business Requirements View (BRV) is based on a simplified subset of REA, with an unclear boundary between the Knowledge Infrastructure and Operational Infrastructure. It is more technically oriented. It does not contain anything to do with agents, which are replaced by Party Types. It does, however, identify a Collaboration that is the formal definition of a dialogue between two or more parties. UMM is, unlike REA, defined using UML, the standard way of formally defining models.

The ebXML Economic Modelling Elements include a useful set of worksheets that make it easy to define a subset of UMM-BRV. The ebXML model is designed to produce an implementable model.

The ECIMF Business Context Equivalence will need to define the types of resources exchanged, the timing of events and the organizations/roles involved. It must preserve the transaction boundaries as these typically involve the legal consequences.

Note: The use of the word Equivalence was identified as being misleading. It was felt that the word Matching would be more relevant.

The ECIMF Business Context Models map to the Economic Exchange View and the Business Process View. The first view defines a set of constraints on exchanges, while the second defines the boundaries between activities as defined by documents exchanged by parties. It defines the Give and Take (or Take and Give) associations as collaborations between agents. These are defined in terms of UML activity diagrams rather than REA fishbone diagrams. In the example given the assessment of needs does not cause events. This is an item for discussion prior to agreement starting, though it might lead to an offer being made.

Methodologies for Ontology-based Semantic Translations – Heiner Stuckenschmidt

Herr Stuckenschmidt explained the work being done under the BUSTER Project by the TZI Intelligent Systems team at the Center for Computing Technologies of the University of Bremen. Looked at the ontologies used for systems integration in projects such as SIMS, OBSERVER, Infosleuth, Ontobroker, SHOE, MECOTA, etc, but not those used for knowledge bases.

BUSTER reviewed the use of the ontologies, the representation used, how ontologies were mapped and engineering aspects such as acquisition support and reuse. Existing systems are often based on single ontologies, a global vocabulary, which may be modularised. This depends on clear boundaries between modules. The alternative is to link sets of local ontologies whose boundaries are not well defined. This means that mappings must deal with the overlap between the systems. A hybrid approach is to have local definitions that use a shared vocabulary to define the terms that appear in more than one ontology.

Global ontologies become difficult to scale up to cover wide ranges of domains. Multiple mappings can only grow by defining more n-n mappings for each new ontology. The amount of work required grows exponentially. The hybrid approach reduces the amount of work required.

Many of the reviewed systems were based on description logics that describe types and their subtypes (e.g. CLASSIC, GRAIL and OIL), but little else. Other work included rule bases that can be used to check relationships (e.g. CARIN, AL-log and DLR). Some systems were based on frame-based representations.

Integration tasks require mapping of ontologies. In some cases this is done by requiring structural resemblance (e.g. tables that define classes), or by defining a set of shared terms. Structural enrichment and meta-annotation (e.g. RDF) can also be used to identify relationships between data sets.

Mapping can be direct (term A to term B) or based on lexical relationships (synonyms, homonyms, etc), a top-level ontology defined in a formal representation. BUSTER tries to create semantic correspondence based on shared terms.

Open questions relate to how to map ontologies formally, rather than the existing ad-hoc mapping between ontologies. The BUSTER architecture separates the Semantic Level, the Structural Level and the Syntax Level. A Mediator is used to map each structure into a "wrapper" that provides a uniform interface to different heterogeneous information sources. Resources are labelled with terms taken from a shared

terminology that can be mapped together. The labels allow correspondences between structures to be computed. This allows mapping rules to be generated. Operators such as AND, OR and NOT, extended by OF (specialization) and COMP (combination) are used to link together terms from the vocabulary that describe a resource. Case-based reasoning and knowledge-based relationships are also used to identify relationships.

The BUSTER ontologies work on 3 levels: level 0 is the operational data in the ontology resources, level 1 defines the types of relationships between the terms while level 2 identifies the shared ontologies used to map terms. They use Upper-Cyc (an ontology of ontologies) and Pangloss as upper-level ontologies, scientific classifications (for plants, chemical substances, etc), domain thesauri (Gemet for GIS terms) and linguistic thesauri (WordNet). They define the relationships in RDF, using OIL as the representation language for display purposes.

Project web page is at www.semantic-translation.com. Slides are at www.tzi.de/~heiner/Methodologies.pdf. Contact heiner@tzi.de.

Knowledge Engineering Tools and Semantic Translation – Andrzej Bialecki, WebGiro

There are different types of tools, based on ontology engineering, knowledge acquisition or automated reasoning (e.g. inference). The same real-world entity can be described by different sets of terms taken from different ontologies that are related to the processes being undertaken by their users. Describing a product in terms of its physical characteristics and the characteristics of the packaging used to ship the product may use similar terms, but the meaning of these terms will be context dependent. Contexts can be used to disambiguate terms. Upper-level ontologies can be used to define contexts.

Protégé 2000 (http://protege.stanford.edu) uses a frame-based knowledge models based on classes and slots (characteristics). It allows for multiple inheritance and templates for slots. The system is modular and extensible. The tool will automatically produce forms for capturing new entries using the designed knowledge model. A visualization unit (OntoViz) can be used to show the ontology as a graph. Anchor-PROMPT allows semantic matches to be defined in terms of matched pairs of anchors. Relationship scores are calculated on number of maps between terms that are linked within the structure. Anchor-PROMPT can, apparently, correctly identify relationships with up to 60% accuracy. Plain PROMPT, which is the only version for which implementation is currently publicly available, only works based on lexical matching. It allows you to merge all parents (from the root), or all subclasses or all instances. While available as open source software with explicit support for semantic matching the underlying knowledge model is difficult for non-experts to understand.

The Conzilla Concept Browser (http://www.conzilla.org) uses a knowledge model, known as the Unified Language Modeling (ULM), based on "neurons" that define classes, specializations of classes (subclasses), associations, aggregations and instantiations. Subclasses define the *isa* relationship as used in knowledge engineering, while instantiations do not map to *isa*. The language has been designed to map easily to natural language. The tool can output both to a proprietary XML data structure and an RDF description.

Conzilla could be used as a plug-in to Protégé, providing an easy-to-use interface to the formal model used for Protégé, which is probably the most mature of the existing ontology mapping tools.

Intelligent Agents and Integration of Business Processes – For Fredric Camuzard

Andrzej Bialecki pointed out some of the key slides from the presentation that Frederic Camuzard of Motorola was going to present. Based on FIPA-conformant agents, the Lightweight Extensible Agent Platform (LEAP) will provide a Java-based platform for the application of generic agent services. Agents will be able to ask questions of other agents based on standardized dialogues (speech acts) based on messages conformant to a "semantic language".

The FIPA Semantics Framework is working on semantics for conventions and contracts (which seem to be similar to ebXML Collaboration Protocols). It includes trust features, ontologies and contract terms.

Note: A break in the presentations was taken between 14:00 and 15:00 to attend a presentation from the European Commission on Framework Programme 6 for IST Research and Technology Development.

Catalogue Integration Methods – Prem Couture, Cyscom

The cXL ontology system developed by Cyscom is used to enrich structured or unstructured catalogue data to allow it to be associated with a knowledge base that is based on a classification tree with a set of properties that can be used to qualify entries. Tools are also provided for creating search filters that are associated with specific classifications.

The Cyscom system allows inheritance of properties from parent classes, with extensions (and restrictions) at any level. Schemas can be customized using a tool provided as an extension to the Excel spreadsheets used to store the catalogue metadata. The system allows the attachment of images to specific fields. Measurements are treated separately as each type of measurement has different relationships with larger and smaller measurements. Classes can have "variations" that create time-limited subclasses of a class that can be used to identify special subclassifications (e.g. free extra contents for special offers, etc).

The system has Creator, Aggregator and Integrator components. The Creator creates new schemas for describing catalogues, the Aggregator can be used to combine catalogues, and the Integrator maps from one schema to another schema.

Co-ordination and reuse

The MULECO project proposal submitted to the EU IST project will seek to develop a Multilingual Upper-Level Electronic Commerce Ontology. Timing is such that the ontology will not be available to be used in ECIMF directly, but it seems obvious that the MULECO ontology would form a useful shared ontology for the ECIMF semantic translation mechanism.

The newly established Open Source Supply Chains project will also provide information that will help in the mapping functions of ECIMF.

The contacts made via ebTWG to the REA team will lead to ECIMF evaluating REA, and providing European feedback on the REA Enterprise Script.

Contacts between the BARBi project team at EPM Technologies and the MULECO project team shows that they have a mechanism for turning STEP AP221 product libraries into displayable multilingual ontologies/classification scheme. Study of this model might provide some useful feedback.

Discussion on project deliverables

While basically on schedule, recent identification of new potential concepts, such as REA, means that significant changes need to be made to the existing documentation. Andrzej needs help in updating the data, particularly in reviewing the data for inconsistencies. Andrzej will take a first cut at updating the text in the next few weeks, for review within both the project group and the wider workshop.

Detailed descriptions of the scenario (currently in a separate document FIG) should be placed in the Proof-of-Concept CWA, with an outline of this in the main guidelines. The possibility of developing some worksheets to provide a short-cut to documenting procedures for creating maps should be considered.

We need to consider possible XML representations (MANIFEST) for the mapping diagrams developed.

The use of XSLT as a constraint and mapping language has been mooted for describing the constraints to be applied to information entities. Andrzej is worried that XSLT, like JavaScript, is not strongly typed. This was a deliberate decision, based on the need for extensibility. Our requirements state that the selected methodology should be strongly typed. Martin Bryan pointed out that you could type the contents of any XML element by checking their content against a pattern at the start of an XSLT statement. (You could use the same technique in JavaScript.) By requiring that pattern checking be a compulsory part of the transformation process you can force the transformation to be strongly typed.

Comments on the deliverables are strongly encouraged, as is new material.

Outstanding Actions

1. Add Business Context concept to meta-model (WebGiro)
2. Find out more about agents in LEAP, and the relevance of ontologies to the use of agents (Motorola)
3. Use of Conzilla as an input tool for developing Protégé models (KTH).
4. Proof-of-Concept deliverable to be based on real-world mapping (WebGiro)
5. Discuss co-operation with EC Workshop's Supply Chains project, and with ebTWG Business Process Collaboration working group.
6. Prepare suggestions for an XML representation of diagrams in the semantic correspondence section.

You can find the slides of the presentations at http://www.ecimf.org/events.html, or in the official CEN/ISSS project document list: http://www.cenorm.be/isss/Workshop/ec/ECIMF/Documents/ECIMF_Documents.htm