Ontology Services-Based Information Integration in Mining Telecom Business Intelligence

Longbing Cao[1], Chao Luo[2], Dan Luo[3], Li Liu[4]

[1,4]Faculty of Information Technology, University of Technology, Sydney, Australia
[2,3]Department of Electronics and Information, Liaoning Technical University, China
{[1]lbcao, [4]liliu}@it.uts.edu.au, {[2]chao.luo, [3]dan.luo}@mail.ia.ac.cn

Abstract. Ordinary implementation of mining telecom business intelligence (BI) is to simply pack data warehouse (DW), OLAP and data mining engines together. In practice, this type of system cannot adapt to changing or new requirements emergent in the problem domain. As a result of survey, 85% of DW projects failed to meet their intended objectives. In this paper, an internal linkage and communication channel, namely an ontology service-based match and translation among user interface, DW, and enterprise information systems, is developed, which implements unified naming and directory of ontology services, metadata management and rule generation for ontology mapping and query parsing among conceptual view, analytical view and physical view from top down. A system prototype on top of realistic telecom environment shows that our intelligence integration solution presents much stronger power to deal with operational decision making user-friendly and adaptively compared with those simply combining BI products available from vendors.

1 Introduction

Mining telecom business intelligence is a complicated process [1]. In this activity, the following procedures must be done: (i) building a DW system on top of Enterprise Information Systems (EIS) running in telecom industry, (ii) supporting four-level data analyses in realistic enterprise environment, and finally (iii) providing and delivering customers of BI system with a flexible and adaptive knowledge portal seamlessly and dynamically, which integrates EIS, DW, OLAP and Data Mining (DM).

As a matter of fact, business requirements of operational analysis and internal structures of underlying EIS are always dynamically realistic. However, almost all ordinary solutions, which have been or are transferred from system integrators of BI to users, are prone to provide users with subjects, data models, analytical dimensions and measures predefined in design time. As a result, most established BI systems cannot adapt to changing or new requirements emergent in the problem domain daily. 85% of DW projects failed to meet their intended objectives, and 40% didn’t even get off the ground [2]. A fundamental reason for the above failure is the poor integration of information distributed among DW, OLAP and DM engines. Simple packing cannot support dynamically analytical requirements and run-time intelligence mining flexibly and adaptively in the real world.

In this paper, we mainly report some of our explorations in constructing an internal linkage and communication channel for information integration from underlying operational systems. We build an ontology services-based integration infrastructure, which implements unified naming, directory and transport of ontology services, and ontology mapping and query parsing among conceptual view, global view and physical view from business interfaces through DW to EIS. This work is from our activities in building telecom BI system by integrating DW, OLAP, DM and reporting systems commercially available. Our experiments in the real world of telecom industry have shown that it can support online and interactive integration of the above mentioned four modules, rather than simply packing together. It can provide users with development supports for adapting to new and dynamic requirements and changes user-friendly and flexibly.

2 Ontology Service Representation

Before going ahead with describing the ontology service-based integration of DW, OLAP and DM, we need to clarify some basic concepts and representations which are essential in defining integration mechanisms based on ontology services[3,4].

Definition 1 Ontology Relationship ($\mathcal{R}$, r): An ontology relationship defines relations existing between two ontologies. An ontology relationship r could be an instance of set of relations $\mathcal{R}$. In our project[5], we define the following elements in $\mathcal{R}$: Identical, Aggregate, Generalize, Substitute, Disjoin, Overlap, and Associate. We further distinguish and relate two or more ontologies with the following predicates: *same_as*, *part_of*, *is_a*, *equal_to*, *disjoin_to*, *overlap_to*, and *relate_to*. Table 1 shows details about ontology relationships.

Table 1. Ontology relationships

Relations	Description	Predicates
Identical	When the two ontologies O1 and O2 are identical	*same_as*
Aggregate	When ontology O2 is part of ontology O1	*part_of*
Generalize	When ontology O2 is a kind of O1	*is_a*
Substitute	When ontology O2 is equal to O1	*equal_to*
Disjoin	When O2 and O1 have no share in common	*disjoin_to*
Overlap	When O2 and O1 have partial share	*overlap_to*
Associate	When O2 is related to O1 in a relation except of the above six	*relate_to or user-defined associating predicates*

Example 1 The following expression shows that ontology Billing consists of LocalBilling and RemoteBilling, while Localbilling and Remotebilling is in Disjoin.

part_of(Billing, *disjoin_to*(LocalBilling, RemoteBilling))

Definition 2 Ontology Organizational Structure Type ($\mathcal{T}$, τ): An ontology organizational structure type τ consists of many ontology elements linked in some relationships. It is a value element of set of types $\mathcal{T}$ which covers all possible ontology organizational structure types. Some main ontology hierarchy, for instance, Hierarchical, Egalitarian, and Hybrid [6,7], can be found in telecom EIS like billing system, operational and maintenance system and so forth.

Definition 3 Ontology Service Item Atom: An ontology service is defined by set of items of attributes and their relationships, which are represented in Key Value Tuples (KVT); all attribute items are embodied and extended from item atoms. An item atom is a basic unit which defines a type of attribute the ontology service must hold, and can be expressed in form as <$\mathcal{A}$>k : v<$\backslash\mathcal{A}$>.

All keys defined in the infrastructure are drawn from a universal namespace, which is extracted and defined by us according to the reality and requirements of analyses in telecom, and encloses complexities in DW, DM and EIS. The pair-element denotes a name in a hierarchical namespace following some type of relationship, where a first token in the tuple is at the highest level in the hierarchy and the rightmost is the leaf.

Definition 4 Ontology Service Description: Following basic item atoms are constituent attributes of an ontology service: <ST>k : v <\ST> is a service type item atom, <SL>k : v<\SL> is a locator item atom, <I>k : v<\I> is an input item atom, <IC>k : v<\IC> is a precondition atom which defines constraints on service, <O>k : v<\O> is an output item atom, the item <IO>k : v<\IO> define constraints across inputs and outputs, <OC>k : v<\OC> is postcondition item of service. Furthermore, each item atom may have a list of items, which is called Item List[8]. For instance, an item list of j inputs is a list of j ($j \geq 1$) items as <I>k1:v1, k2:v2, …, kj:vj<\I>.

For each item atom in a service, there may be some mandatory items and other optional ones. Multiplicity of constraint properties of items can be found in our case, like MO: Mandatory One, *MM*: Mandatory Many, *OO*: Optional One, *OM*: Optional Many.

Example 2 The following is service description for the DMAlgorithmRegistration ontology, which registers a globally unique decision tree algorithm with three mandatory inputs and one optional algorithm creating date. All inputs and outputs of the algorithm are separated by semicolons and stored into array in Java, and generates one new decision tree class if registration is successful. A token /$\mathcal{O}$ is marked at the end of all optional items.

<OS> register: algoid(datamining)
<ST>Type: DecisionTree<\ST>
<SL>LocationID: ioas. algorithm.datamining.algoid<\SL>
<I>Name:String,InputString:StringArray,OutputString:StringArray,Date: Calender/$\mathcal{O}$*<\I>*
<IC>InputString: semicolon, OutputString: semicolon<\IC>
<O>Algo_class: DecisionTreeClass<\O>
<OC>Status: successful<\OC>

<IO>Relation: one-one<\IO>
<\OS>

Accordingly, we can define some basic ontology services (as partially shown in Table 2) needed in the integration of DW, OLAP, DM and reporting system in telecom industry.

Table 2. Partial ontology services

Ontology Service	Verb	Noun-term
SubjectAddition	create	subjectid(serviceroot_subject)
DimensionUpdate	update	dimid(serviceroot_dimension)
MeasureDeletion	delete	measureid(serviceroot_measure)
DMAlgorithmRegistration	register	algoid(algorithmtype)
DMAlgorithmUpdate	update	algoid(algorithmtype)
DMAlgorithmExecution	execute	algoid(algorithmtype)
OntologySearch	search	ontologyid(serviceroot_ontology)

3 Ontology Integration of Business, Data Warehouse and EIS

Our major objective of heterogeneous information translation and integration includes: (i) providing transparent and seamless integration of the underlying heterogeneous resources among telecom operational systems, (ii) supporting smooth transformation from business concepts in user interface to low-level entities in specific resource systems, and (iii) furnishing online interactive techniques for transparent interoperability and smooth translation among levels.

3.1 Ontology Match and Translation Structure

As an approach for overcoming heterogeneity, ontology has been widely investigated and used for explicit description of information in heterogeneous resources. In summary, there are three ways of deploying ontologies: single ontology approach, multiple ontology approach and hybrid ontology approach[6]. It is believed that hybrid ontology approach can overcome drawbacks of single and multiple ontology approaches [6, 7]. However, there is still lot of research work on how to build a hybrid ontology infrastructure for integrating domain specific heterogeneous resources, from business to DW to underlying EIS.

In the process of integrating heterogeneous and distributed information from data sources, we proposed an ontology structure as shown in Figure 1. There are three views coexisted in this ontology system from top down: a top-level conceptual view, an analytical view, and a low-level EIS view. A mediator level is built for ontology mapping and query parsing among three levels. A universal namespace, a representation of ontology services, and ontology mapping and query parsing are introduced for uniquely naming, resolving, identifying and transporting ontologies and their relations among levels. In addition, KVT/ Key Value Pairs (KVP) and Key Property Pairs (KPP, here Property is used to define cardinality of an entity) are utilized for describing ontology naming, directory, location and transport services.

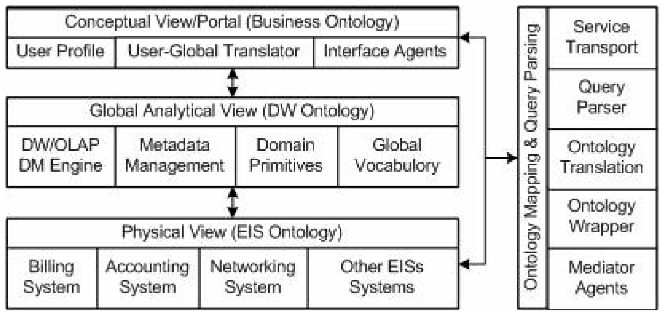

Fig. 1. Ontology match and translation

3.2 Top-Level Conceptual View

The objective of Conceptual View is to present users with domain specific concepts, objects, business rules, and user interfaces in a conceptual profile user-friendly. The ontologies capture general knowledge about concepts, terminology and relationships from viewpoint of business in the world.

The output of this view is a conceptual ontology base, which includes a Concept Category Directory (CCD). The CCD, which is a hierarchical concept tree implementing telecom business namespace, lists and defines all terms and relationships abstracted in daily business, and generates a list of candidate concepts and expressions based on the business process and activities happened in the user views. Here, a concept rather than an attribute or entity is used to describe the world. For instance, *Conditions…* rather than *Where…* is used in generating a query.

Definition 5 Concept Category Directory Entry: A CCD entry consists of a unique Leading Item (LI, an identifier), and optionally multiple Substitute Items (SI, recommended candidate concepts) as follows:

{{{*Leading Item, MO*}:{*LI_Value*}},{{*Substitute Items, OM*}:{*SI_Value1, SI_Value2,…* }}},

for instance:

{{{*LI_Service_Provider*}:{*Service_Provider_Label*}},{{*SI_Service_Provider*}:{*Service_Provider_Name,Service_Provider_Nickname,Service_Provider_Description* }}}.

3.3 Global Analytical View

The Global Analytical View is a logical aggregated representation of underlying logical elements and relationships locating in DWs, OLAP server and DM engines. So, in terms of domain specific primitives in telecom information systems, this global side ontology wraps technical and business metadata items. These metadata items are defined in set of elements (attributes, dimensions and measures) in data model, source data, ETL, and also actions and rules of interaction between DW and data sources.

Definition 6 Analytical Ontology Directory Entry: It consists of some domain-specific metadata items, which focuses on business and technical metadata required in the problem domain, rather than on business rules and concepts. The following elements are enclosed in it: globally unique identifier(gui), recommended global name(rgn), candidate substitute names(csn), parent object(po, top-level coupled concept name), child objects (co, low-level EIS instances), analytical locator(al, where to find this entry from the bottom EIS resources, including related connection string, schema, metadata of resources, and so forth), close associators (ca, including actions and relationships with other neighboring entries). Furthermore, the cardinality property of an entry is shown in the following:

{{gui, MO}, {rgn, MO}, {csn, OM}, {po, MO}, {co, MM}, {al, MO}, {ca, OM}}

All item atom entries are stored into knowledge base and registered into ontology name database. The usage of KVT, KPP, the Pair-Element encoding system, and the introduction of elements *parent* and *child, locator* and *associators,* can solve conflicts of data type, scaling, generalization, naming and location.

Furthermore, in order to build this orderly aggregated analytical level, a metadata management mechanism is necessary and built-in for organizing data model of DW, OLAP and DM engines. It also encloses information about bottom resource primitives, data connections, resource locations, and services and queries distributions.

3.4 Low-Level Physical View Ontology

The Low-level Physical View is a representation of physical entities and relationships related to transactions among underlying information systems. The most common form of Physical View is as tables and attributes located in EIS. EIS enclose multiple enterprise information resources in which store huge amount of operational data and information. On this level, telecom operational systems like BOSS, MIS, ERP, OA are all resource providers of the DW and DM system.

In terms of technical implementation, the multiplicity of this level also brings us a colorful world of physical instances/attributes/relations and so forth. For instance, the counterpoints of *Customer_Name* on user conceptual view, may take names as *Customer_Name, Customer_Label, User_Name, User_Label* etc. in physical systems(Figure 2). These names may be distributed into the following business operational systems like billing system, accounting system, switch system, operation and maintenance system, etc.

4 Ontology Mapping & Query Parsing

The objective of this level is to construct a set of mediation services which implement online ontology mapping and management among top-level concepts, global metadata items and low-level resource attributes. It also implements query parsing from concept-oriented query requests in user interfaces to DW elements, EIS attributes and relations-based query statement in the bottom.

4.1 Ontology Mapping among User View, Global View and Physical View

Ontology mapping first implements match between user view in concepts and global view in elements and relationships, then transports user requests from global to low-level physical view in concrete attributes existing in information resources.

Figure 2 shows principle we cope with the above three-level match. As described above, three-level views are separated for specific requirements of different levels. The above ontology representation is used to organize elements on each level.

As mentioned before, for each record of conceptual ontology, the unique Leading Item (actor) may have 0 to many Substitute Items (stand-ins). For instance, a leading item of *Customer_Name* may have many stand-ins, for instance, *User_Name, User_Label,Customer_Label*, and so forth. After fixing the mapping from concepts to DW elements and EIS attributes, owners of BI system can define or update itemsets as they have or like by interfaces.

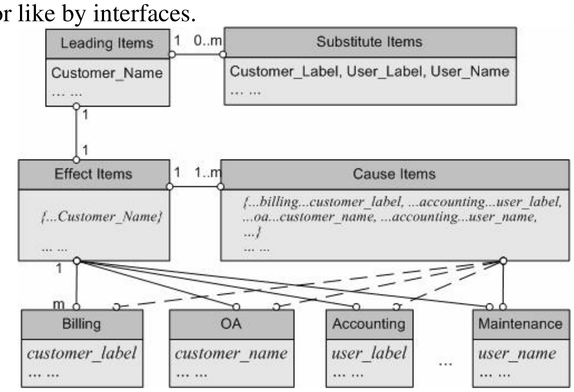

Fig. 2. Ontology mapping

In the global analytical view, there are many Cause-Effect pairs there. For each element of global ontology, there is only one Effect Item, which uniquely identifies a metadata item of global view on the basis of DW models. This Effect Item wraps related element in data model of DW and data marts. On the other side, one to many concrete instances/attributes/relations compose the corresponding Cause Items (as shown in the following pair elements), which are exactly some specific instances, attributes, what we would see in respective EIS and data sources on the low level.

{{Effect Item, MO}, {Cause Items, OM}}

It is noteworthy that here all naming of ontology elements on the high level and the physical level are in the same namespace, level-oriented ontology directories, and based on the same serviceroot. Moreover, an ontology name database is built to manage the naming and labeling, resolving and match of logical name and physical name of ontologies on different levels. A user interface module is developed to deal with activities of Insert, Delete, Modify and Query the logical and physical names of the ontologies in the name database.

4.2 Query Parsing Process

Query processing is defined by parsing rules. A query parsing rule can be generated automatically by selecting metadata items, defining relations (by predefined predicates) among ontology and metadata items, setting limit conditions (constraints) for relevant items and attributes, and grouping and sorting where appropriate.

An interactive module (as shown in Figure 3) is developed to support online definition and generation of query rules on the basis of metadata management. This module not only arranges the translation and integration among conceptual level, DW level and physical level, but also covers metadata items of underlying source data, for instance, ETL rules, and mapping from EIS to concepts directly.

Here, the Relations toolbar is in charge of ontology and metadata relationship management. Limit Condition is used to set constraint property among items. Grouping and Sorting are for arranging items in group or sorting in ascending or descending directions.

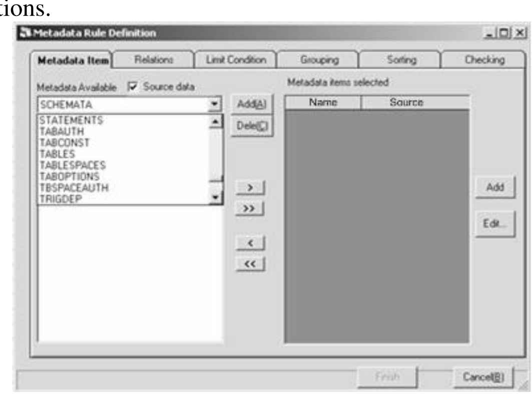

Fig. 3. Interactive query rule generation

As a prerequisite, some modules are developed for construction, registration and maintenance of ontologies and metadata items. Predicates of ontology relationships can be invoked from user interface of ontology construction to define relations among ontologies. Ontology naming and index are managed from ontology registration module and further stored into ontology name database and knowledge base.

5 Case Study: a Prototype for Intelligence Integration in Telecom

According to what we have discussed in the above sections, we have constructed a business intelligence prototype system called IOAS[5]. It has integrated DW server, OLAP server, DM engine on top of realistic business and data in mobile telecom, and formed a unified knowledge portal for enterprise decision making. Figure 5 presents some screen shots of this system.

In the IOAS, IBM DB2 Universal database is used as database server, DB2 Data Warehouse and Oracle OLAP server stores six subjects and six special subjects, Cognos is used for OLAP, ad hoc and predefined reporting presentation, IBM Intelligent Miner mines business intelligence from huge amount of telecom data. Under the IOAS, Informix, Oracle, Sybase, DB2 are used in telecom business operational systems for storage of respective operational transactions.

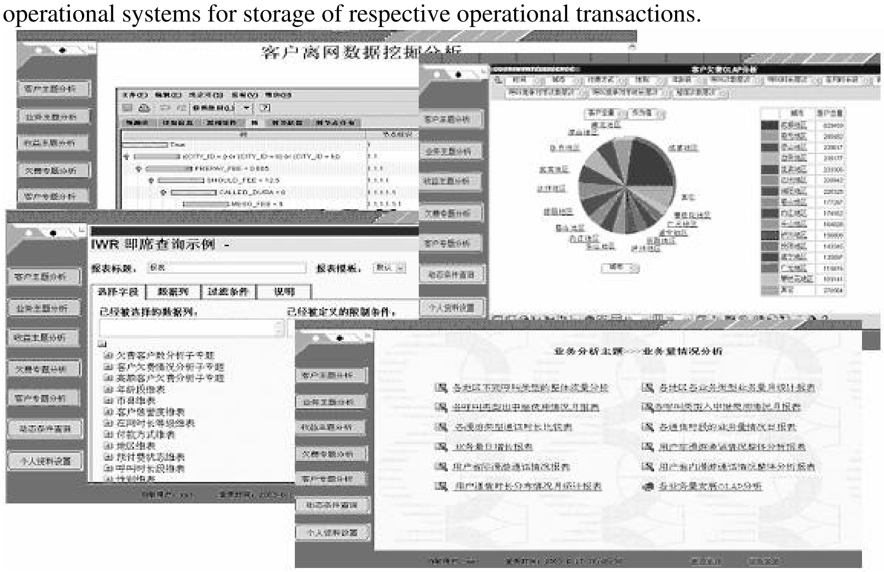

Fig. 4 Case study: IOAS integrating business intelligence in telecom

To users, complexities from multiple heterogeneous information resources, ETL tools, Operational Data Store system, DW system, presentation reports and tools, are shielded and hidden under the one-stop interface. They can easily launch analysis and observations without worries of underlying heterogeneity, symbolization, authorization and information management from DW to bottom data sources.

6 Performance Evaluation

Compared with solutions simply combining BI products presently available from vendors, our ontology services-based approach presents much stronger power for information integration user-friendly and flexibly from the following aspects:

(1) The integration of analyses, DW, OLAP and DM is not by simple addition on the basis of a reporting system, rather through an internal ontology service-based linkage and communication channel. This three-level hybrid ontology schema sets up some internal mechanisms for supporting ontology mapping, query parsing among business concepts, analysis models and physical entities.

(2) Some user-friendly development supports help users to modify, update, create or re-arrange ontologies and functionalities on different levels in terms of problem domain and requirements. Thus, users can arrange their own three-level ontology base and ontology namespace as required adaptively without modifying the

match relations. Again, the metadata management tool helps users to maintain and match ontology mapping and query parsing in terms of their interests.

(3) Representation and directory of ontology services, combined with others like metadata rule definition, support structural and semantic transparency in dealing with the heterogeneity and interoperability among domain specific levels.

Technically, our solution by providing internal location, directory, mediation and transport supports can help BI system users enhance their own capabilities in dealing with changing or new environment flexibly and adaptively in a user-friendly manner.

7 Conclusions and Future Works

In this paper, we studied how to integrate information among reporting, DW, OLAP and DM engines beyond realistic telecom business operational systems. We introduced a three-tier ontology infrastructure implementing transparency from business concepts to underlying EIS. Uniform namespace, ontology representation, and metadata management are used to define, locate and transport ontology elements among three views. An ontology mapping is designed to transparently associate relevant ontologies with Conceptual, Global and Physical Views. Interactive rule generation supports dynamic information translation and query parsing among low-level attributes, metadata items, and concepts-based business profiles.

We further built a BI system prototype on the basis of realistic mobile telecom operational systems and historical data. It has shown that it is more user-friendly, flexible and adaptive for telecom customers to online mine business intelligence from huge amount of business transactions, than simple packing those BI components.

Our future works include but are not limited to the follows:

(1) Representation and deployment of FIPA-compatible ontology services;

(2) Increasing adaptability and run-time working power to deal with dynamically evolutionary business environment and decision-making requirements.

References

[1] Mark Whitehorn, *Business intelligence: the IBM solution*, New Springer, 1999.

[2] Usama M. Fayyad. Tutorial report of Summer school of data mining. Monash Uni, 2003.

[3] FIPA. FIPA0006 and FIPA0086.

[4] Hiroki Suguri et al. Implementation of FIPA Ontology Service.

[5] Longbing Cao, Dan Luo, Chao Luo, Chengqi Zhang. Systematic engineering in designing architecture of telecommunications business intelligence system. *Proceedings of HIS'03*, 1084-1093, IOS press, 2003.

[6] H. Wache, et al. Ontology-based integration of information – a survey of existing approaches. 2001. From webster.cs.uga.edu/~budak/ courses/s02/nis/Survey-2.ppt.

[7] Zhang cui, Dean Jones and Paul O/Brien. Issues in ontology-based information integration. www.csd.abdn.ac.uk/ebiweb/papers/cui.pdf.

[8] Cheng Hian Goh. *Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Sources*. Phd, MIT, 1997.