

Develop the Ontology for Internet Commerce by Reusing Existing Standards

Yuxiao ZHAO

Dept of Computer and Information Science, Linköping University, S-58183, Linköping, SWEDEN

yuxzh@ida.liu.se

<http://www.ida.liu.se/~yuxzh/>

Abstract

This paper attempts to present an approach to develop the ontology for Internet commerce by reusing the vocabulary from the existing standards, industry initiatives or XML-based frameworks. The approach allows ontology developers to share the common vocabulary and build it from a solid foundation. The paper includes two main contributions: 1. an approach to develop the ontology for the domain of Internet commerce; 2. the guidelines about what vocabulary can be reused from the existing standards.

1. Introduction

Ontology is the key enabler in the technology of semantic Web (SW). To develop applications of SW for Internet commerce, we will first have to build the well-crafted ontology¹ for the domain to facilitate SW adoption (Euzenat 2002). The ontology demands at least two requirements:

1. A formal ontology language is necessary, for instance, Web ontology language (OWL), DAML+OIL (McGuinness et al 2002), XML Schema or RDF². W3C is developing OWL, and its success will significantly ease the problem.
2. It needs to foster interoperability, *Ontological commitment is important. It is the agreement by multiple parties (person and software systems) to adopt a particular ontology when communicating about the domain of interest, even though they do not necessarily have the same experiences, theories, or prescriptions about that domain* (Holsapple and Joshi 2002).

¹ The ontology in this paper indicates the ontology for the domain of Internet commerce.

² Gil and Radnagar (2002) compared three kinds of semantic markup languages, XML, RDF (Resource Description Framework) and DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer). For practical reasons their conclusion is to use XML Schema for developing user-oriented tools operating over the semantic Web.

To meet the second requirement, the ontology engineers should probably consider the reuse. There are two types of ontology reuses, one being direct reuse from existing ontology library³, and the other being reuse from existing vocabulary⁴. Either way will save a great deal of efforts and time compared to building the ontology from scratch.

We attempt to provide an approach to develop the ontology for Internet commerce by reusing existing industry standards, or XML-based frameworks focusing on the integration issues in Internet commerce (Zhao 2001). Further, we try to answer what vocabulary can be reused and how to reuse them. This paper is organized as follows: Section 2 describes the approach for the ontology development. Section 3 addresses how to capture the vocabulary and how to reuse them. Section 4 concludes the paper and identifies the future research.

2. An approach to develop the ontology for Internet commerce

This section is to present the approach for developing the ontology by reusing XML-based standards. Figure 1 shows the processes to develop the ontology for the Internet commerce:

- **Applications** in the top box represent the programs we are developing.
- **Ontology Requirements** are to serve for the applications operating in an ontology-sharing community.
- Based on the ontology requirements, **Vocabulary Capture** is a method for collecting the set of related concepts by reusing the standards, which may include standards such as XML Schema (XSD), RDF/RDFS (RDF Schema), ISO (International Organization for Standards) standards, ANSI (America National

³ Such as DAML ontology library:
<http://www.daml.org/ontologies/>.

⁴ Ontology in the semantic Web is the formal specification of concepts and relationships between them. The vocabulary is the set of the concepts.

Standards Institute) standards, and UBL (Universal Business Language), xCBL (XML Common Business Language), OAGIS (Open Applications Group Interface Specification), and RosettaNet. In the next section the process is further discussed in depth.

- **Ontology Enhancements** are to modify some concepts and axioms to meet customized requirements and add some other requirements that we have not covered but of which the applications do demand. So we demand some ontology tools like Protégé-2000 (Noy et al 2001), which support the ontology language we prefer.
- Of course, the produced **Ontology Documents** should be validated and verified to show whether it meets the applications requirements.

The approach reuses existing XML-based standards and it has the advantage of leveraging a variety of best practice and expert knowledge that the standards depend on, for instance, naming and design conventions, the form of the concepts, relationship modeling, taxonomy, type definitions. Compared to building the ontology from scratch, we could save a great deal of time and money but produce the ontology with higher interoperability that will probably be needed in the future. The advantage will become increasingly obvious with the growth of the ontology size.

Since we will determine the extent to which we reuse the standards at the start of the ontology development, the concern about reuse will not overly influence our choice of ontology tools and formal development processes. We may for example choose nothing in the start and then come back from the following processes.

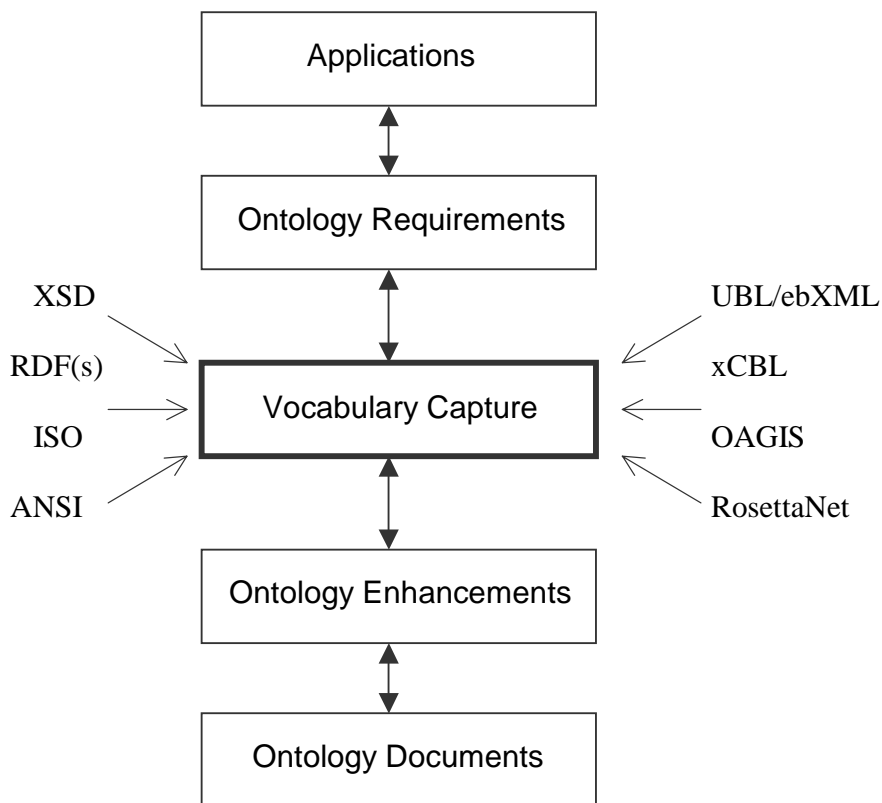


Figure 1: The processes of developing the ontology for Internet commerce

3. Vocabulary Capture

3.1 Available standards to be used

Our researches have realized that there exist many industry initiatives, or XML-based frameworks for Internet commerce, which are really valuable for building the ontology. Zhao and Sandahl (2000) compared those major frameworks that were active at that time, i.e., Microsoft's BizTalk, Ariba's cXML (Commerce XML), Commerce.net's eCo Framework, IDEAlliance's ICE (Information and Content Exchange) protocol, IOTP (Internet Open Trading Protocol) in IETF (Internet Engineering Task Force), Open Application Group's OAGIS, RosettaNet's PIPs (Partner Interface Processes), and CommerceOne's xCBL. Zhao (2001) further addressed other eight emerging frameworks focusing on B2B e-business, i.e., ebXML (e-business XML), UDDI (Universal Definition, Discovery, and Integration), SOAP (Simple Object Access Protocol), WSDL (Web Services Description Language), RDF, OFX (Open Financial eXchange), VoiceXML, and Wf-XML (Workflow XML).

Our research demonstrated that all frameworks involve different types of vocabulary that we try to reuse in the ontology development. Table 1 shows these frameworks' functionality, which is needed in the development of Internet commerce. Last column in Table 1 denotes the emphasis of vocabulary in the frameworks. We can find seven frameworks have heavy vocabulary involvement, i.e., ebXML, cXML, ICE, OAGIS, RosettaNet, UBL and xCBL. Our latest research indicates that four frameworks tend to have higher possibility of contributing to the ontology, i.e., xCBL, UBL (Universal Business Library), RosettaNet and OAGIS (see Table 2). We drop ICE protocol because it focuses on information syndication and drop cXML because it gives more proprietary solutions for its e-procurement software and its influence on the reuse is relatively low. OASIS (The Organization for the Advancement of Structured Information Standards) initiated UBL project later than other frameworks but some characteristics make it substantially important in the future, e.g., its neutrality, takeover of UDDI, offering BIEs (Business Information Entities) for ebXML, as well as its basis on existing frameworks like xCBL.

Table 1: XML-based frameworks and their functionality (Zhao 2001)

Name	Registry	Search	Contracting	Configuration	Messaging	Syntax & Semantics	Vocabulary
ebXML	X	X	X	X	X	X	X
eCo Framework	X	X				X	x
UDDI	X	X		X		x	x
SOAP					X	x	x
BizTalk					X	x	x
cXML					X	X	X
ICE					X	X	X
OAGIS					X	x	X
RosettaNet					X	x	X
OFX					X	x	x
VoiceXML					X	x	x
RDF						X	x
UBL						X	X
WSDL						X	x
Wf-XML						X	x
xCBL							X

X---major concentration

X--- involved

Table 2: Vocabulary in four industry initiatives

Initiative	xCBL	UBL	RosettaNet	OAGIS
Hyperlink	www.xcbl.org/ xCBL 4.0	www.oasis-open.org/committees/ubl/	www.rosettanel.org Business Dictionary 2.1 Technical Dictionary (RNTD 2.0)	www.openapplciations.org Open Applications Group, OAGIS 8.0
Players	CommerceOne, Inc.	OASIS ebXML	RosettaNet Consortium, a subsidiary of UCC (Universal Code Council)	Most members are vendors of ERP, EAI and SCM
Goals	To develop XML documents and framework to allow interoperability by exchanging robust, reusable, XML documents. It essentially serves as the "mother code," to let all participants can understand.	To modify an existing XML library (xCBL) for ebXML Core Components and to design a mechanism to generate context-specific business schemas by XPath.	Aims at integration between supply chain partners, transaction consistency and the elimination of communications inefficiencies in electronic component, information technology and semiconductor manufacturing companies.	to define a unifying standard for e-Business and Application interoperability. It is the largest publisher of XML based content for business software interoperability in the world.
Focus	Define common data types for e-procurement	Develop the standard library and the mechanism	Define PIPs (Partner Interface processes) and related dictionary	Defines BODs (Business Object Document) & models any interaction as a VERB and NOUN
Vocabulary Type	XSD (XML Schema)	XSD; UML (Unified Modeling Language); Spreadsheet	XSD	XSD; XSLT (XML Stylesheet Language Transformation); XPath
Vocabulary Inclusion	Core; Catalogue; Application integration; Order management; Preorder management; Financial; Material management; Message management; Statistics and forecasting.	Core library: UBL Trading Cycle Will include: Vendor managed inventory; Self-billing; Master Order and Calloffs; Prior Quote Request & Quotation; International Trade requiring Multi-party; Transportation; Hire Trade.	Technical dictionary (www.rntd.info/draft/): product classification Business dictionary 2.1: BusinessDataEntities; BusinessProperties; FundamentalBusinessDataEntities	CRM (Customer Relationship Management); Contracts; Core Components; HRXML (Human Resource XML); Product Data Collaboration; Logistics; Warranty.
Form as example of order	SimpleType: OrderTypeCodeType PaymentTermCodeType RequestedResponseCodeType ServiceCodeType ComplexType: OrderTypeType PaymentTermType RequestedResponseType ServiceLevelType	UBL Trading Cycle: Order Order Response (simple) Order Response (complex) Order Cancellation Dispatch Advice Receipt Advice Invoice	BusinessDataEntities: OrderConfirmation OrderForecast OrderForecastItem OrderInformation OrderLeadTime OrderNumberInformation OrderQuantity OrderReference OrderShippingInformation OrderShippingInstruction OrderStatus OrderStatusQuantity BusinessProperties: OrderCountry OrderDateTime OrderForecastQuantity FundamentalBusinessDataEntities: OrderForecastQuantityTypeCode	Abstracts: Verbs (Acknowledge, Get, Post, etc), Nouns (PurchaseOrder, Credit, Receipt, etc). BODs , e.g. AcknowledgePurchaseOrder. Many scenarios (e.g. 12.0 Purchase Order Process) are defined by aspects of scenarios diagram, assumption, component definition, business workflow, exception handling.

3.2 Reuse is not so simple

Reusing the standards is theoretically not difficult. For example, we can use some query language and logic combinations such as *intersection* and *union* to search what we want. Unfortunately, they are not available in databases; instead, they

- Reside in distributed Web sites setting different access policy. For OAGIS, we have to register first, download the zipped file, unzip it, and read the unzipped files via a browser. For xCBL, we don't need to register personal information but we directly access to them via its Web site. For UBL, there is no formal library for the vocabulary, instead there are many documents to discuss the mechanism and methodology how to create the vocabulary, showing it is in the initial stage¹. For RosettaNet, you have to accept the license conditions first, then download-unzip-read the unzipped files locally.
- Use different data/document modeling, i.e., xCBL uses a pragmatic way to abstract the documents for e-procurement software development. UBL is trying to apply document engineering and data-centric way and categorize any document as content, structure and presentation, considering xCBL and standards from ISO and UN for ebXML Core Components. RosettaNet considers product classification in technical dictionary besides business entities and their properties. OAGIS models any integration scenario as the combination of a *verb* and a *noun* such as *GetReceipt*.
- Use different schema (types and constraints of files) to specify their vocabulary and/or relationships, although XML Schema has been unanimously used.
- UBL is still in initial stage while other three kinds of vocabulary are updating frequently. We showed the latest version of the frameworks in Table 2.

Therefore, the problem has actually been transformed into knowledge merging and acquisition, that is, querying multiple, access-constrained, form-differed and distributed Web sources, which can be challenging. This research is not to solve this problem but to compare and contrast the vocabulary in these frameworks, categorize, analyze and

¹ For the latest UBL version, 0p70, published on Jan 27, 2003, there is a core library. To access to it, you have to download the zipped file, unzip it and read them via a browser-this could be a general pattern for accessing to such a huge source.

structure them, and then present the layered vocabulary for reuse.

3.3 Layered vocabulary for reuse

We find that some vocabulary represents general business information, documents, activities, processes and interactions, and thus has higher use frequency, while other vocabulary is specialized and not often used. Layered vocabulary reflects the classification of the vocabulary. Figure 2 shows the layers. The dark circle signifies **Core** vocabulary that probably is the most frequently used in Internet commerce applications. **Core** vocabulary may reuse some **General** vocabulary, which is very general for applications of various fields. **Reusable** vocabulary represents some relatively specialized community, but it also can be reuse in these communities. **Special** vocabulary is extremely customized for one community based on Special and/or Core vocabulary.

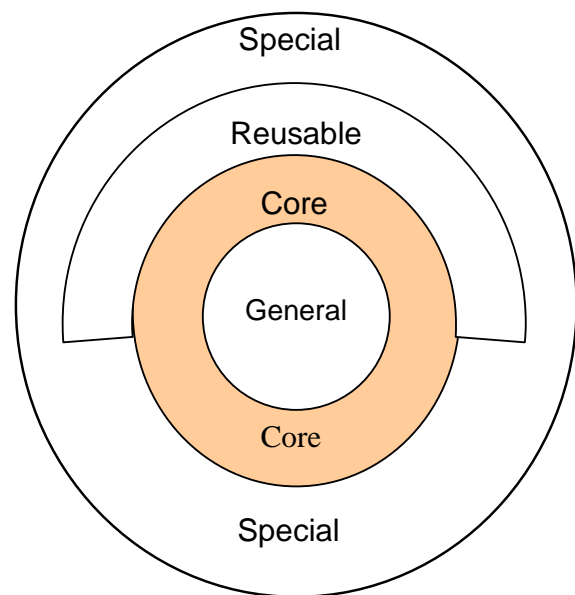


Figure 2: The layers and relationships of the common vocabulary for Internet commerce

As we understand the layers, we go further to answer two important questions: (1) what contents should be included in each layer? And (2) which frameworks can be reused for them?

Our research shows that besides XML, XML Schema and RDF/RDFs, which can be directly used, the following

components should be addressed in **General** vocabulary, which will be reused more broadly:

- Date and Time: ISO 8601; RosettaNet; OAGIS
- Currency: ISO 4217; OAGIS
- Geography such as Country: ISO3166; OAGIS
- Language (speaking): ISO639
- Quantity: RosettaNet; OAGIS
- Address: Name and Address Markup Language (NAML); RosettaNet; OAGIS

Core vocabulary should include:

- Stakeholder (ownership, manager): RosettaNet
- Organization (type, manager, middle manager, personnel): RosettaNet; OAGIS
- Human Resource management: OAGIS; RosettaNet
- Service: RosettaNet for traditional services; DAML-S (DSC 2002) for Web services
- Catalog (price, product, database, etc): UBL; xCBL, RosettaNet; OAGIS
- Material Management (inventory, hardware, software, equipment, part, component, maintenance, etc): xCBL; UBL; RosettaNet; OAGIS
- Procurement (PurchaseOrder or PO, PO request, PO response, PO change, etc): UBL; xCBL; RosettaNet; OAGIS; cXML
- Payment (invoice, credit, confirmation, deadline, security, etc): xCBL; UBL; RosettaNet; OAGIS
- Accounting (account, ledger reports, etc): xCBL, OAGIS; RosettaNet
- Marketing/sales (brand, advertising, customer, user, etc): OAGIS; RosettaNet
- Product data: RosettaNet; OAGIS; UN/SPSC¹ offers standardized products and services codes using five layers of Segment, Family, Class, Commodity and Business Function
- Customer data: xCBL; UBL; OAGIS; RosettaNet
- Statistics: UBL; xCBL; RosettaNet; OAGIS
- Security
- Privacy

Reusable vocabulary should include:

- Transport (status request, shipping, etc): UBL; OAGIS; RosettaNet
- Contracting (form, signature, execution, evaluation, etc): RosettaNet; OAGIS

- Project management (goal, budget, personnel, schedule, status, change, process, etc): OAGIS
- Middlemen (subcontractor, service provider, etc): RosettaNet; OAGIS
- Quality (ISO standards, certificate, etc): ISO; RosettaNet
- Industry: RosettaNet offers technical dictionary in industry of electronic component, IT and semiconductor; many industry initiatives coexist².

Special vocabulary can be built on the basis of the common vocabulary and can be extended in terms of:

- Geographic characteristic (climate, political, legal, habitual differences)
- Context (merger, spin-off, financial conditions)
- Business strategy

4. Conclusions and future work

This paper provides an approach to develop the ontology for Internet commerce by reusing XML-based standards or industry initiatives. We address nearly all influential initiatives that we can reuse for Internet commerce, in particular, four standards, UBL, xCBL, OAGIS and RosettaNet, will become very important. We also categorize the vocabulary into four layers and related standards for easy reuse. The four layers are called General, Core, Reusable and Special.

For future research, we plan to identify all **Core** vocabulary to build the **Core** ontology using OWL or DAML+OIL for practical use. Then we test the ontology by developing some applications for Internet commerce in order to evaluate the advantages of SW in this domain. Zhao and Sandahl (2003) identified several applications such as semantic routing. In addition, we consider designing some models and implementing some tools to ease the standards reuse.

Acknowledgements

The author would like to acknowledge Kristian Sandahl, Angela Yong and two anonymous reviewers for valuable comments.

This work has been funded by SSF (Swedish Strategic Foundations) through ECSEL (Excellence Center in Computer Science and Systems Engineering), Linköping University, Sweden.

¹ See <http://www.unspsc.org/>

² See <http://xml.coverpages.org/xml.html#applications>

References

- DSC: DAML Services Coalition (alphabetically A. Ankolenkar, M. Burstein, J.R. Hobbs, O. Lassila, D.L. Martin, D. McDermott, S.A. McIlraith, S. Narayanan, M. Paolucci, T.R. Payne and K. Sycara), 2002. DAML-S: Web Service Description for the Semantic Web, *International Semantic Web Conference (ISWC2002)*, Sardinia, Italy.
- Euzenat, J., 2002, Research challenges and perspectives of the semantic Web, *IEEE Intelligent Systems*, 17(5), pp86-88.
- Gil, Y. and V. Ratnakar, 2002. *Markup language: comparison and examples*, available <http://trellis.semanticweb.org/expect/web/semanticweb/comparison.html> (current Dec 5, 2002).
- Holsapple, C.W. and K.D. Joshi, 2002. A collaborative approach to ontology design, *Communications of the ACM*, 45(2), pp42-47.
- McGuinness, D.L., Fikes, R., Hendler, J. and L.A. Stein, 2002, DAML+OIL, *IEEE Intelligent Systems*, 17(5), pp72-80.
- Noy, N.F, Sintek, M, Decker, S., Crubezy, M, Ferguson, R. W., and M.A. Musen, 2001. Creating semantic Web contents with Protégé-2000, *IEEE Intelligent Systems*, 16(2), pp60-71.
- Zhao, Y., 2001. *XML-based frameworks for Internet commerce and an implementation of B2B e-procurement*, Licentiate's Thesis No 882, LiU-Tek-Lic-2001:19, Linköping University Electronic Press, http://www.ep.liu.se/lic/science_technology/08/82/index.html (current Dec 1, 2002).
- Zhao, Y. and K. Sandahl, 2000. XML-based frameworks for Internet commerce, *Proceedings of International Conference on Enterprise Information Systems (ICEIS2000)*, UK, pp511-516 (also in *Enterprise Information Systems II*, Kluwer Academic Publishers, 2001, pp197-202).
- Zhao, Y. and K. Sandahl, 2003. Potential advantage of semantic Web for Internet commerce, to appear in *Proceedings of International Conference on Enterprise Information Systems (ICEIS2003)*, Angers, France, April 23-26, 2003.