Utilizing Ontologies in eCommerce

Avigdor Gal
Rutgers University

© Utilizing Ontologies in eCommerce, by Avigdor Gal

eCommerce

- Electronic Commerce (eCommerce) is the process of two or more parties making business transactions via computer and some type of network.
- The business practices and technologies known collectively as eCommerce include:
  - Using email for business communication
  - Using Electronic Funds Transfer (EFT) with bank, suppliers, and clients
  - Establishing a Website to market products or services
  - Implementing an Internet-based retail business
  - Implementing Electronic Data Interchange (EDI) with suppliers and clients
  - Integrating EDI with in-house financial systems
  - Implementing Technical Data Interchange (TDI) to exchange eng. drawings
  - Implementing compatible eCommerce practices throughout acquisition to distribution supply network
  - Creating a “virtual enterprise”

http://www.ecrc.gmu.edu/ecommerce/intro.html
eCommerce

- Typical classification:
  - Business-to-Business
  - Business-to-Public Body
  - Business-to-Individual (22% of eCommerce and declining rapidly)
  - Individual-to-Public Body

B2B figures: (based on Jupiter Comm.)
- 2000: $336 Million
  3% of transactions
- 2005: $6 Trillion
  42% of transactions

* 22% from consumer transactions 78% from business-to-business transactions.
** 8.8% out of the revenue from consumer transactions and 91.2% from business-to-business transactions.

By 2002

Sources: Yankee Group, IDC, Forrester Research
West Chester Electronic Commerce Resource Center

B2B eCommerce

- Any commercial transaction involves both information and physical exchange between parties (merchant, customer).
  - information exchange: product information, transaction information,…
  - physical exchange: product, cash, service,…
- eCommerce represents the trend of decoupling information exchange from physical exchange for such transactions, and the delivery of the former through IT.
- eCommerce: islands (of automation) in the sun.
Data Integration in B2B eCommerce

• For B2B eCommerce to work, IT has to deliver on **data integration**: fetching, integrating (syntactically and semantically), and presenting data from distributed, heterogeneous information sources.

• Data integration is intended to meet information needs for a community of users, based on the contents of one or more information sources.

• The quality of a data integration service is measured by the soundness and completeness of the information provided, as well as by the quality and speed of presentation.

• Data integration is a key in the successful utilization of ontologies in B2B eCommerce.

Utilizing Ontologies in eCommerce

• Tutorial goals:
  – Review existing approaches towards utilization of ontologies.
  – Identify the unique needs of eCommerce, with respect to data integration.
  – Discuss the role of research in this field.

• Audience:
  – **Practitioners** interested in obtaining the data management perspective of this ever-growing field.
  – **Researchers and students** interested in doing research in ontologies and eCommerce.
Presenter

Avigdor Gal is a faculty member at the Department of MSIS at Rutgers University. He received his D.Sc. degree from the Technion-Israel Institute of Technology in 1995 in the area of temporal active databases. He has published more than 30 papers in journals (e.g. IEEE Transactions on Knowledge and Data Engineering), books (Temporal Databases: Research and Practice) and conferences (e.g. ER'95, CoopIS'98) on the topics of information systems architectures, active databases and temporal databases. Together with Dr. John Mylopoulos, Avigdor has chaired the “Distributed Heterogeneous Information Services” workshop at HICSS’98 and he was the guest editor of a special issue by the same name in the International Journal of Cooperative Information Systems. Also, he was the General co-Chair of CoopIS’2000. Avigdor has consulted in the area of eCommerce and is a member of the ACM and the IEEE computer society.

Thanks

- John Mylopoulos
- Vijay Atluri
- Brad Kingsburg
- Suren Krishnamurthy
- Manju Radhakrishnan
- Venkatesan Esakki
Outline

- Introduction
- The role of ontologies in eCommerce
- Existing ontology tools for eCommerce: examples
- A toolkit for data integration
- Research Issues

The role of ontologies in eCommerce
Disclaimer: what is an ontology

- Webster: the branch of metaphysics dealing with the nature of being.
- A narrow definition: vocabulary [SMITH99]. An ontology offers a standard vocabulary for a given subject matter, say manufacturing, medicine, banking or retailing.
- My personal choice: concepts related through relationships.
- Examples:
  - Relational schema (foreign keys)
  - XML’s DTD (more about it later).
  - Semantic nets.

ontology.org

- ontology.org is an independent industry and research forum, formed in 1998, that is focused upon the application of ontologies in Internet commerce.
- According to the director of ontology.org “[t]he key ingredients that make up an ontology are a vocabulary of basic terms and a precise specification of what those terms mean.”
Electronic supply chain

- Inter-enterprise activities:
  - Purchase order
  - Advance ship notice
  - Invoice
  - Electronic funds transfer
- An ontology enables automatic processing of the information.

Electronic market exchange

- An electronic exchange place is a meeting point for agents to exchange information and goods.
- Each agent uses an autonomous data source (may it be a database or XML-based documents).
- To facilitate information exchange, each agent is also equipped with an ontology, e.g., a database catalog or an XML's DTD -- Data Type Definition).
**Information services**

- The role of information:
  the capability of deciding among several actions which are otherwise not distinguishable [WOODWARD55].
- Ideal Information:
  » good quality
  » compact representation
- The information crisis:
  “drowning in information... but starved for knowledge.” John Naisbitt, Megatrends.
- An information service is any service (e.g. computerized activity) that is designed to provide compact, coherent and reliable information obtained from distributed, heterogeneous information sources on an as-needed basis.
- The coherency of information depends on having an ontology for mapping the data from the heterogeneous sources.

**A compiled list of references on ontologies**

- [http://wiscinfo.doit.wisc.edu/is/ontology/](http://wiscinfo.doit.wisc.edu/is/ontology/)
- Includes references to works by:
  - Y. Wand et. al., UBC, Canada.
  - N. Guarino et. al., CNR, Italy
  - H. Kim et. al., (TOVE), York U., Canada
  - Others…
Existing ontology tools for eCommerce: examples

Electronic Data Interchange (EDI)

- EDI is a standard of computer to computer interchange of business documents in a machine-readable form.
- Leading EDI standards: EDIFACT [BERGE91] (Europe) and ANSI X12 [HAMILTON93] (North-America).
- A common EDI scenario:
  - A customer sends an EDI purchase order to a provider (transaction 850).
  - The provider returns a purchase order acknowledgment to the customer (transaction 855).
  - At ship time, the provider sends the customer an EDI advance ship notice (ASN) stating the goods to be shipped (transaction 856).
  - Once the goods have been shipped and billing has been processed, the provider sends the customer an EDI invoice (transaction 810).
  - The customer sends payment information to its bank, which wires funds to the provider’s bank through electronic funds transfer (EFT) or financial EDI.
EDI (semi) structure

- Components of an EDI message:
  - data element: a datum identified by a reference number. Examples: date, price, purchase order number, quantity on order.
  - data segment (or: segments): a sequence of data elements. For example, a purchase order segment consists of part number, part description, quantity, unit of measure and cost.
  - transaction set (or message): a set of data segments. For example, an invoice (ASC X12 set number 810).
  - functional group: a set of similar transactions.
  - Syntax rules: specifying the grammar for EDI dialogue.

Expressing EDI messages

- EDI can be expressed using OEM [PAPAKONSTANTINOU95].
- Each functional group is represented as a graph:
  - internal nodes are complex values (e.g., transaction sets).
  - leaf nodes are atomic values of data elements.
  - edges represent semantic relationships and are labeled for naming purposes.
- An example, based on [ADAM98]:
**eXtended Markup Language (XML)**

- XML is a tagging language, in which tags can be defined by the user, rather than by the vendor, as is the case with HTML.
- Common ways of defining XML tags (based on [SAHUGET2000]):
  - DTD, Document Type Definition. Enables:
    - Parsing
    - validation
  - XML schema: a W3C product, which constrains the meaning, usage, and relationships of the XML constituent parts. [W3C2000]
  - Type checking: validation of a DTD. [WALLACE99]
  - Constraint based: the structure of an XML document is defined using constraints and pattern matching. [JELLIFFE99]
  - UML. [KIMBER99]

**BizTalk**

- BizTalk (http://www.biztalk.org):
  - “BizTalk is an industry initiative started by Microsoft and supported by a wide range of organizations, from technology vendors like SAP and CommerceOne to technology users like Boeing and BP/Amoco.”
  - “BizTalk is not a standards body. Instead, we are a community of standards users, with the goal of driving the rapid, consistent adoption of XML to enable electronic commerce and application integration.”
XML.org

- “XML.ORG is a credible, independent resource for news, education, and information about the application of XML in industrial and commercial settings.”
- Hosted by OASIS, the Organization for the Advancement of Structured Information Standards. OASIS is a nonprofit, international consortium dedicated to accelerating the adoption of product-independent formats based on public standards.
- “Through XML.ORG, OASIS will collect, manage, and distribute information about XML applications, including vocabularies, schemas, namespaces and DTDs.”

A toolkit for data integration
A toolkit for data integration

- Translator:
  - wrapper
  - specification language that is:
    - common
    - expressive

- Integrator:
  - semantic reconciliation
  - semi-automatic/automatic process

- Coordinator:
  - run-time interpreter
  - scheduler

Example 1: Data integration at the Library of Congress*

- Translator: Descriptive cataloging, e.g., Dewey Decimal Classification.

- Integrator: Cataloging Policy and Support Office.

- Coordinator: Librarians

* The source of information and the picture is http://lcweb.loc.gov/catdir/
Example 2: Data integration in Federated databases

- Translator: Common Data Model (CDM).
- Integrator: schema integration [BATINI86]
- Coordinator: basically DDBMS functionality

Translator: Wrappers

- Wrapping an information source amounts to defining an interface, i.e., what data can be passed in/out, and a protocol, i.e., what methods are supported by the information source for accessing or updating its contents.
- For example, a Web wrapper needs to support the following protocol: In response to a connect method, the wrapper returns an object with the following methods:
  - Set of input attributes of basic types that can be converted to strings; these accept bindings for a query;
  - Set of output types, can be tuples or complex objects;
  - An input-output relationship which is allowed to use set, bag and other constructors;
  - Optionally, the connect call may handle meta-information (is the wrapper relevant to the query, how much will the query evaluation cost...)

[SHETH90]
Example: Weather Information

Suppose you have various URLs from which you can get weather information (say, 3-5 day forecasts) for different European or international cities, e.g.,

http://www.weather.com/weather/us/cities/MD_Baltimore.html, or
http://www.weather.com/weather/int/cities/ES_Barcelona.html

We’d like to be able to retrieve weather information for any city
We need to write software which constructs URLs, also extracts automatically data from an HTML document.

Translator: Specification languages

• The specification language should be as expressive as the underlying data source’s language.

• Specification languages:
  – Data-centered:
    • SQL
    • XML
    • Semantic nets [HULL87]
    • HTML
    • IDFT WebSQL [WEBSQL]
  – Process-centered:
    • Workflows
**Integrator:**

Semantic reconciliation

Semantic reconciliation is called for whenever there is a disagreement about the meaning, interpretation, or intended use of the same or related data.

---

**Data Repositories**

- A **repository** is a generic information system intended to represent, store, manage, disseminate and maintain data **about data** (computer-based, but also non-computerized information sources)
- A repository is like a library index, it tells you where to go to find the information you are looking for; repositories store **metadata**, (i.e., data about data.)
- A repository consists of an Information Model, an Information Base, an Information Base Management System, and Control Functions.
- A repository becomes handy in managing ontologies.
Commercial Repository Products

- Repositories products are bundled with a number of products in the market.
- Total repository market is estimated to be ~$1B/yr.
- Repositories are being used for a variety of applications:
  - For software development -- Andersen Consulting, IBM, ORACLE, Softlab, Sterling, ...
  - For information resource management -- Platinum, Viasoft, MSP, Unisys
  - For application management -- SAP, Baan, PeopleSoft, ...
  - For product data, Web/doc and system management
  - For data warehousing
- Why aren’t repositories the primary databases for the data they describe? Because of performance, human factors, ...

Repository Architecture

Information Model
- predefined types;
- highly extensible

Repository Engine
- objects, properties;
- rich relationships;
- versioning, ...

DBMS, relational or OO
**Integrator: semi-automatic/automatic process**

- **The problem:**
  - building ontologies requires manual tuning
  - initial task is time consuming.
  - change detection and propagation is needed to ensure currency.

- **The solution:**
  - machine assistance for building ontologies.
  - growing ontologies rather than building them.
  - automatic change detection.
  - semi-automatic propagation.

**Integrator: Machine assistance**

- Data mining
- **IR:**
  - Extraction of structure in business reports [LIDDLE99].
  - Using existing ontologies to identify new ones [LABROU99].
**Integrator: growing ontologies**

- Dynamic Classificational Ontologies [KAHNG96]:
  - Two types of ontologies:
    - (Base) static: requires the agreement of the parties.
    - Dynamic: may evolve dynamically.
      - based on ownership.
      - Two forms of growth:
        - export
        - discovery

**Integrator: change detection and propagation**

[Image of characters: WebMonitor, Coordinator, Designer, ConceptEditor]

[GAL99]
WebMonitor

- WebCQ
  - A change tracking system for Web pages.
  - Monitor various types of changes to static and dynamic Web pages.
  - Personalized delivery of page change notifications.
  - http://www.cc.gatech.edu/projects/disl/WebCQ/
- CoopWARE’s WebMonitor
  - Monitors a single URL or a set of URLs.
  - detects and informs a user of any insertions, modifications and removals of Web artifacts within its scope.
  - Supported by a Mini SQL 2.0 database.
  - WebSQL is utilized to query the information under the conditions specified by a user through the interface
  - Changes are sent to user by eMail.
  - http://business.rutgers.edu/~webmon

Coordinators

- We are going to look at three types of experimental architectures for coordination:
  - Mediators;
  - Task-oriented coordination;
  - Reactive coordination
Coordination by mediation

- Mediator: an intermediary that performs retrieval, filtering, transformation and presentation functions on data.
- Related issues:
  - standardizing the structure of mediators;
  - establishing a common set of mediator functions;
  - setting a standard for mediator communication.

Sample Mediation Architecture
Research efforts in mediation

- TSIMMIS at Stanford University: http://www-db.stanford.edu/tsimmis/tsimmis.html
- MIT's Context Interchange project: http://context.mit.edu/~coin/
- SIMS information mediator at USC: http://www.isi.edu/sims/
- The Cooperative Database Project at UCLA: http://www.isi.edu/sims/

Task-Oriented coordination

- Data integration is accomplished in a context defined by some activity. For example,
  - Task 1: customer fetches book prices in order to buy, or
  - Task 2: a merchant fetches book prices in order to set her own book prices
- Workflows can be used to specify tasks.
- The OZ project adopts this perspective to data integration [http://www.psl.cs.columbia.edu/fact-sheets/oz-fact-sheet.html]
- OZ supports a rule language. A rule generally corresponds to a workflow step.
- Rules can be executed in a forward- or back-chaining mode
Reactive coordination

- Data integration is a coordinated activity among a number of data sources and other components for purposes of retrieval or update (like a meta-mediator).
- Coordination is accomplished through event-condition-action rules.
- Every data source, or other component of the architecture can provide services and signal events, which trigger rules that are executed on the part of a central coordinator.
- This architecture is particularly suitable for control integration among data sources, i.e., if something happens here, something else must happen there…
- The CoopWARE project explores this research direction.

The CoopWARE Architecture

[Diagram showing the CoopWARE architecture with data sources, interface libraries, rule sets, rule engines, and knowledge bases (KB) schemas.]

[GAL99a]
Research Issues

© Utilizing Ontologies in eCommerce, by Avigdor Gal

Research Challenges for Repository Technology

- Version control:
  - Need a mechanism for representing and naming versions;
  - Operations for deriving a new version from given ones;
  - Semantics for checkout and checkin;
  - Constraints on version histories;
- Configuration management
  - Representations for configurations;
  - Operations on configurations, such as attaching or detaching a component;
  - Semantics of change propagation; if a component is changed, is the whole configuration changed as well?
- Context management
- Tool integration
Interoperability

- Generic, open architectures.
- Distributed object management.
- Agent-based or network-centric computing. [PAPAZOGLOU92]
- Compartmentalized applications.
- Factoring out global control from individual components.
- Communication protocols
- Translation mechanisms

[MYLOPOULOS97]

Knowledge Integration:
DARPA’s Knowledge Sharing Project

- An initiative to develop the technical infrastructure to support the sharing of knowledge among systems.
- Four major activities within the initiative:
  - An interlingua for knowledge interchange -- in order to address the different languages problem
  - Standardized semantics for a class of languages, -- say terminological languages
  - Communication protocols for knowledge bases -- situated above system communication protocols
  - Development of generic, reusable knowledge bases

[PATIL92]
Coordination

- Computer supported collaborative work
- Synchronous and asynchronous sharing
- Virtual workspaces, performers and customers
- Concurrency control
- Multi-agent systems and technologies
- Transaction management
- Mediation architectures
- Workflow systems
- AI planning

[MYLOPOULOS97]

Change management

- Changes dictated by technology and/or organizational objectives.
- Constraint enforcement
- Schema evolution
- Database view updates
- AI theories of action
- Truth maintenance systems
- Constraint satisfaction
- Versions and configurations
- Impact analysis
- Risk assessment
- Business process reengineering
- Enterprise integration.

[MYLOPOULOS97]
System management issues

• Extranets:
  – Relational transducers [ABITEBOUL98]
  – Obsolescent data [GAL99b]

• Inter-Enterprise Workflow Models:
  – Trust [MINSKY98]
  – Architecture [GAL99c]

Bibliography

Bibliography