

ECCMA 11TH ANNUAL DATA QUALITY CONFERENCE

- ONTOLOGIES FOR E-COMMERCE STANDARDS

- Donald Hillman

Department of Computer Science and Engineering
Lehigh University

djh3@lehigh.edu

October 13, 2010

Outline

- Goal
- Approach
- Feasibility Study
- Conclusions

Goal

- To develop a user-friendly tool for creating consistent structured ERP descriptions for electronic business transactions
- To provide automatic assistance for creating contracts, RFQs, and purchase orders

Structured Data Descriptions

- Every business area has its special vocabulary and set of concepts
- It is necessary to standardize communications in these areas to provide efficient transactions
- The need is greater with the increased use of electronic commerce

Pursuit of Rigor in Data Standards

- Old-Style: ISO 14258 – concepts and rules for industrial automation enterprise models
- Data Model Standards: ISO 10303-41 STEP standard for the exchange of product model data

Semantic Models Standards

- Cyc: comprehensive system of everyday commonsense knowledge – goal to perform human-like reasoning – MCC project
- ebXML: OASIS/CEFACT
- IEEE 1900.1; SUMO suggested upper merged ontology – set of structured concepts upon which specific domain ontologies can be constructed

Tool-Building Approach

- Use the ECCMA Open Technical Dictionary (eOTD) as the reference architecture for electronic business applications
- This embodies the semantically-based approach now prevalent in the pursuit of data rigor and quality

ISO 22745

- ISO 22745 provides a standard for specifying the syntax, semantics, and data requirements for transactions involving master data exchanged between organizations in the supply chain

The eOTD

- The ECCMA Open Technical Dictionary is the first ISO 22745-compliant dictionary
- It contains terms, definitions, and images linked to public-domain identifiers
- These are used to create unambiguous language-independent characterizations of entities

THE eOTD

- The eOTD employs a form of semantic encoding compliant with ISO 8000110:2008 for the exchange of quality master data
- Companies may use the eOTD to create cataloging templates or Information Guides (IGs), to formulate data requests, and to measure data quality

Poor Data Quality

- Tell-tale symptom: “free text” descriptions in ERP systems – inconsistent, unverified, and unverifiable declarations
- S. de Jager “Solving Data Quality Problems Using Data Standards”, Logistics Spectrum, Oct-Dec 2007
- G. Radack, “ISO 22745: The Standard for Master Data”, **ibid.**

Data Purification

- Application of data quality standards
- Conversion of free-text to structured master data
- Use of structured master data to create structured descriptions

Data Purification Project

- Project (2007 – 2009) to purify different classifications by mapping them to the Federal Catalog System (FCS) – now the NATO Codification System
- FCS developed and maintained by the Defense Logistics Information Service (DLIS) of the Defense Logistics Agency
- DLIS needed to consider the catalogs of private sector vendors in procuring products

Data Purification Project

- Motivation
 - Heterogeneity underlying customers' product, catalog, and document descriptions
- Taxonomies and classification schemes
 - More than 40 have been publicly identified
 - eCI@ss, UNSPSC, CPV, NAICS, etc.
- Federal Catalog System
 - Naming, classifying and describing items of supply for DoD – created & maintained by DLIS
 - Various codes: FSG, FSC, INC, NSN, MRC

Characteristics of the FCS

- The FCS is the original basis of the eOTD
- It has > 400,000 Approved Item Names, with associated attributes, units of measure, currencies, enumerated values, etc.
- It is a taxonomy, with Federal Supply Groups (FSGs) and Federal Supply Classes (FSCs)

Feasibility Study

- Because of the similarity between the eOTD and the FCS, the DLIS Project constitutes a proof of the feasibility of automatic eOTD mapping
- The project was completed on time and under budget
- The results were presented at the International Semantic Web Conference, October 2009

Project Information

TAXONOMY MAPPING USING ONTOLOGY STRUCTURES

Contract Number: SP4701-07-C-0006
Sponsor: Defense Logistics Information Service
Contractor: Lehigh University
Subcontractors: Electronic Commerce Code Management
Association (ECCMA)
Concurrent Technologies Corporation

Effective Dates: September 7, 2007 - March 6, 2009

Project Staff

- Lehigh University
 - Donald Hillman, Professor & Director
 - Jeff Heflin, Associate Professor & Associate Director
 - Basuki Setio, Graduate Research Assistant (9/07 – present)
 - Yang Yu, Graduate Research Assistant (6/08 – present)
 - Ameet Chitnis, Graduate Research Assistant, (12/08)
 - Xinlei Wu, Graduate Research Assistant (9/07 – 1/08)

- Concurrent Technologies Corp.
 - Gerry Radack, Principal Computer Scientist

- ECCMA
 - Peter Benson, President

Problem Addressed



- ◆ Relate Government/Commercial Item Descriptions/Taxonomies to Supplier Capabilities
 - DLIS must be able to integrate DOD item classifications with commercial systems
 - Commercial cataloging systems have domain-specific classifications that are not interconnected
 - Information is stored in different, often proprietary, and incompatible formats

Approach: Guiding Principle



- ◆ Use the eOTD as the basis for the integration of classifications
- ◆ The eOTD is a more comprehensive industry version of the FCS and has recognition as the international standard for e-catalogs via the ISO 22745 designation

Approach: Mapping Limitations



- ◆ Mapping keywords is insufficient
 - one-to-one correspondences aren't always possible
 - overlapping classes
 - » functional vs. compositional classifications
 - » e.g. grinding machine:cutlery (0161-1#01-007071#1) vs. grinding machine:carbide tool bit (0161-1#01-007098#1)
- ◆ Need deeper analysis
- ◆ Focus on systemic classification of attributes that connects communities of information
- ◆ This connection can be established with the Semantic Web

Approach: The Semantic Web



◆ Definition

- The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.
(Berners-Lee et al., Scientific American, May 2001)

◆ Key International Standards

- World Wide Web Consortium (W3C) recommendations
 - » Resource Description Framework (RDF)
 - » Web Ontology Language (OWL)
- Same status as HTML / XML

Approach: Ontology



- ◆ Taxonomy/Hierarchy
 - Classification scheme
 - FSGs, FSCs, and AINs form a taxonomy
- ◆ Ontology
 - A logical theory that accounts for the intended meaning of a formal vocabulary (Guarino 98)
 - Has a formal syntax and unambiguous semantics
 - Can be used to derive logical consequences
 - Usually more than just a hierarchy / taxonomy
- ◆ IGs define informal ontologies
 - Include properties, constraints, valid values, etc.

Approach: Problems



- ◆ Ontologies define strict **is-a** hierarchies
 - $A \text{ rdfs:subClassOf } B$ is interpreted as “Every A is a B ”
- ◆ Many product taxonomies are not is-a hierarchies
 - They were created to support purchasing
 - eCl@ss example:

27 Electric engineering, automation, process control
engineering

27-05 Accumulator, battery

27-05-01 Station. batt., accum.

27-05-02 Traction battery, starter battery

27-05-04 Portable battery

27-05-06 Battery charger

27-05-90 Accumulator, battery (other)

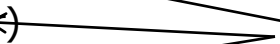
27-05-91 Accumulator, battery (parts)

27-05-92 Accumulator, battery (accessories)

27-05-98 Accumulator, battery (maintenance, service)

27-05-99 Accumulator, battery (repair)

These are not batteries!



Approach: OWL



- ◆ **Web Ontology Language**
 - data is linked to ontologies that provide semantics
 - syntax based on Resource Description Framework (RDF)
 - » but richer semantics
 - » can describe more complicated relationships
 - semantics based on Description logic (DL)
- ◆ **Description logics**
 - form of knowledge representation
 - Provide a rich language for formally defining classes
 - mature reasoning software (FaCT, RACER, etc.)
- ◆ **Benefits**
 - optimized computation of subsumption
 - » calculate implicit subClassOf relations
 - ontology integration
 - » if two ontologies use class expressions to define their vocabularies in terms of a third ontology, then subsumption can be used to compute an integrated ontology

OWL Class Constructors



Constructor	DL Syntax	Example
intersectionOf	$C1 \cap C2$	GasTurbine \cap AircraftPart
unionOf	$C1 \cup C2$	Door \cup Airframe \cup TailSection
complementOf	$\neg C$	\neg Aircraft
oneOf	$\{x1, \dots, x2\}$	{F15, F16}
allValuesFrom	$\forall P.C$	\forall partOf.Airframe
someValuesFrom	$\exists P.C$	\exists hasPart.Door
maxCardinality	$\leq nP$	≤ 10 hasPart
minCardinality	$\geq nP$	≥ 2 hasPart

OWL Axioms



Axiom	DL Syntax	Example
subClassOf	$C_1 \subseteq C_2$	Aircraft \subseteq System \cap Aerospace
equivalentClass	$C_1 \equiv C_2$	Helicopter \equiv Aircraft \cap RotaryWing
disjointWith	$\neg C_2$	RotaryWing \subseteq \neg FixedWing
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	{washer} \equiv {shim}
differentFrom	$\neg\{x_2\}$	{bank:Blood} \subseteq \neg {bank:Capacitor}
subPropertyOf	P_2	hasMaximumWeight \subseteq hasWeight
equivalentProperty	$P_1 \equiv P_2$	weight \equiv displacement
inverseOf	$P_1 \equiv P_2^{-1}$	partOf \equiv hasPart ⁻¹
transitiveProperty	P^+	predecessor ⁺ \subseteq predecessor
functionalProperty	$T \leq 1P$	$T \leq 1$ hasNSN
inverseFunctionalProperty	$T \leq 1P^{-1}$	$T \leq 1$ hasPartNumber ⁻¹

Project Scope



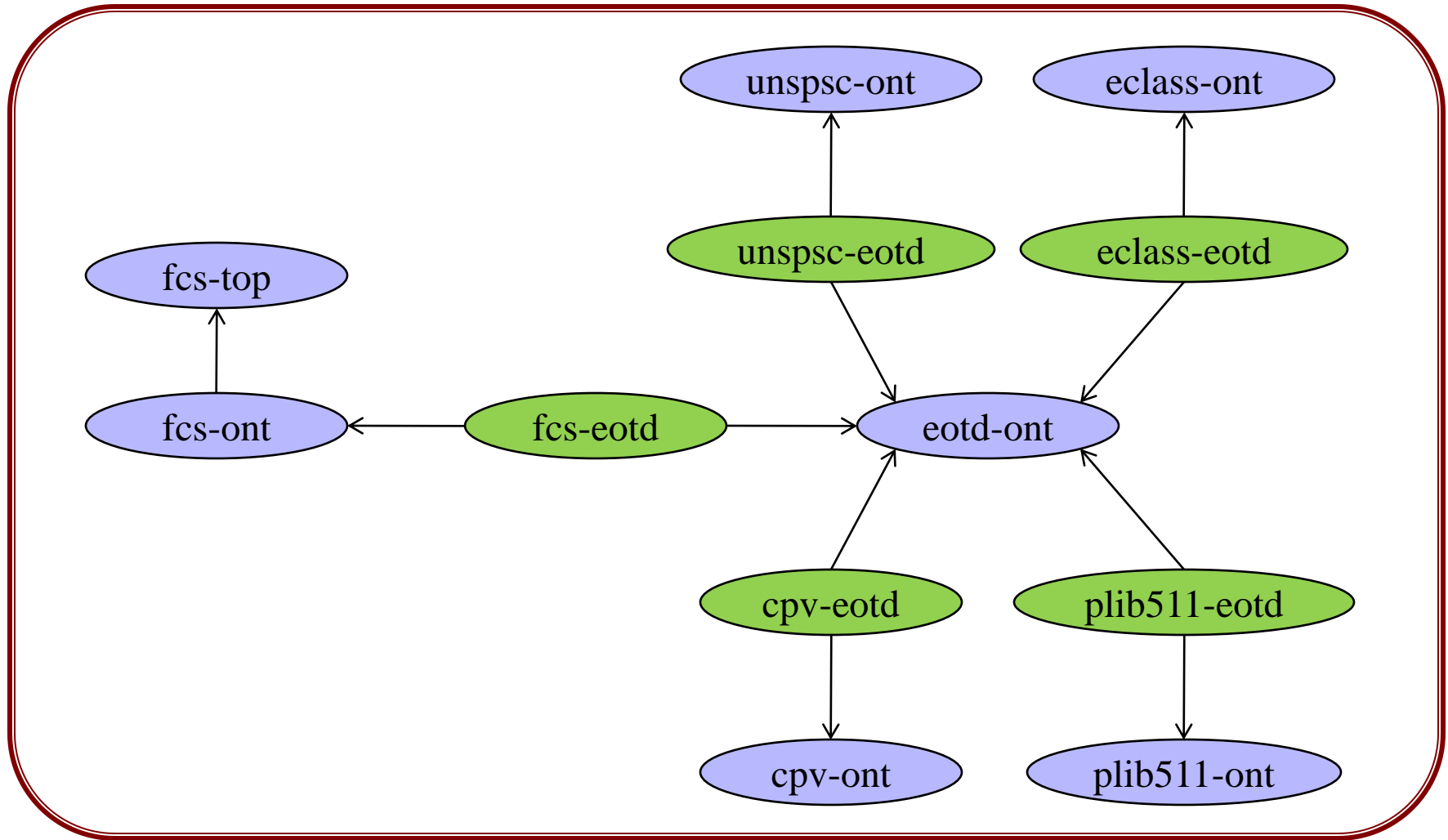
- ◆ Selected Domains for Taxonomy Mapping
 - DLIS specified a set of domains for the project consisting of:
 - » 71 INCs/AINs
 - ◆ selected Batteries, Bearings, Bushings, Fasteners, Gaskets and Microcircuits
 - » 17 FIIGs
 - » 14 FSCs applicable to 490,825 NSNs
 - later added FIIG A003B (BOLTS AND SCREWS) for 57 more INCs/AINs
- ◆ Selected taxonomies
 - United Nations Standard Products and Services Code (UNSPSC)
 - Standardized Material and Service Classification (eCl@ss)
 - Common Procurement Vocabulary (CPV)
 - ISO 13584-511 Fastener Taxonomy
 - » Replaced North American Industry Classification System (NAICS) at DLIS request

Project Stages



- ◆ With the chosen domain and the target taxonomies:
- ◆ Created an OWL ontology for the eOTD
- ◆ Created an OWL ontology for the FCS
- ◆ Created OWL ontologies for the target taxonomies
- ◆ Demonstrated ontology integration
- ◆ Developed ontology translation compiler

Results: Ontologies Overview



Results: Ontology Mapping



- ◆ OWL axioms were used to map each ontology to the eOTD
 - $\text{fsc:BearingsPlainUnmounted} \equiv \text{eotd:BearingPlain} \cap \text{eotd:BearingUnmounted}$
 - $\text{fsc:BushingsRingsShimsAndSpacers} \equiv \text{eotd:Bushing} \cup \text{eotd:Ring} \cup \text{eotd:Shim} \cup \text{eotd:Spacer}$

Results: Ontology Details

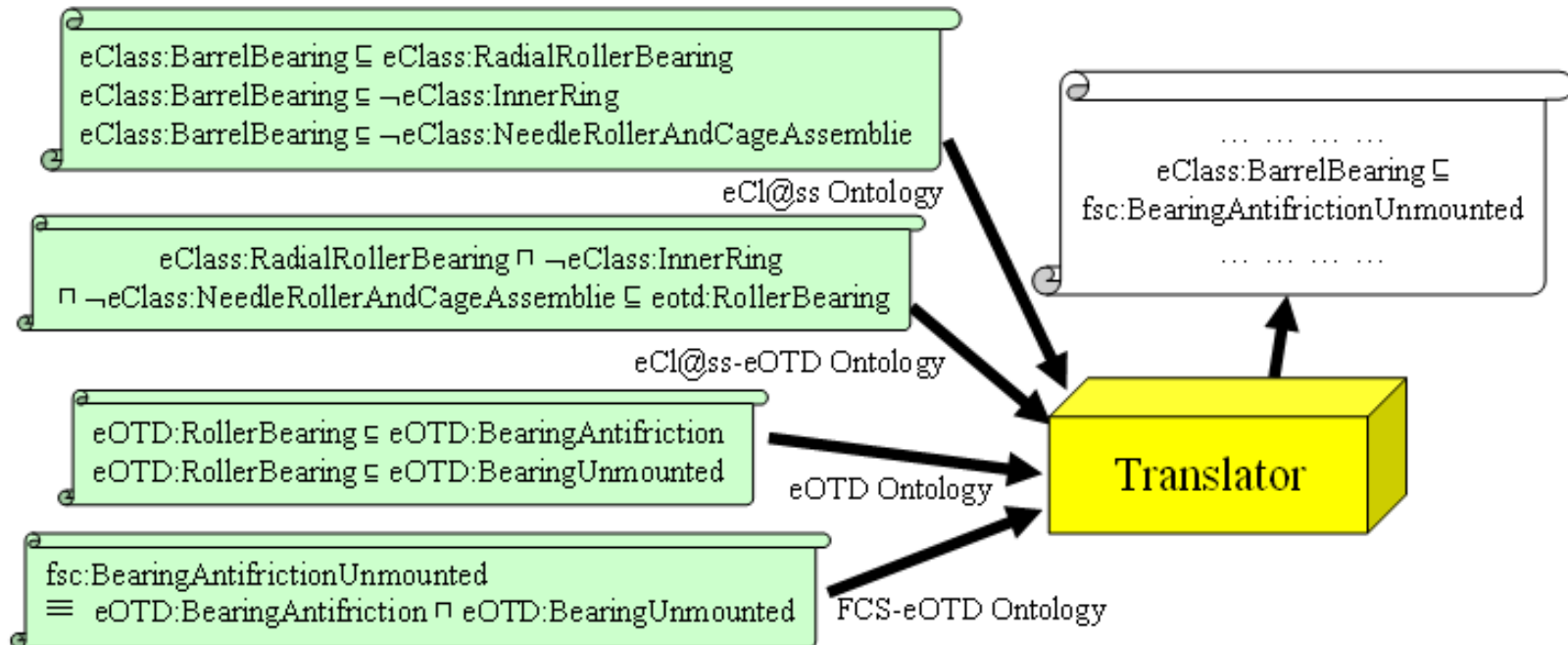


Ontology	Number of Classes	Number of Properties	Depth of Hierarchy	Comments
fcs-top / fcs-ont	819	157	3	FSGs, FSCs, selected AINs, FIIG A003B AINs
eotd-ont	195	191	up to 4*	Corresponds to AINs from FCS. We added some taxonomy
eclass-ont	313	20	4	Original taxonomy has many more concepts and properties.
cpv-ont	208	0	up to 7	Original taxonomy has 3224 concepts.
unspsc-ont	228	0	4	Original taxonomy has 21536 concepts.
plib511-ont	187	199	5+ **	ISO 13584-511: Fasteners standard using PLIB data model. All concepts/properties included.



Implementation

◆ An example of translation



Translation Compiler



- ◆ Automatically generates translation code
 - Input: source and target ontology
 - Output: translator that classifies source data according to target taxonomy
 - » Translation involves deducing relationships between source and target ontologies
 - » In some cases, the nature of the source classification precludes a translation at the AIN level
 - uses DL reasoning to determine relationships
- ◆ Compiler can be rerun any time source / target ontologies change
 - Infrequent, so speed isn't critical
- ◆ Translator is fast
 - Can be run as frequently as data changes

Conclusions



- ◆ The project demonstrated the feasibility of a semi-automatic tool for translating external catalog items into FCS-compliant descriptions
- ◆ The outcome supports the possibility of doing the same thing for eOTD-compliant descriptions

Expected Benefits of Tool



- ◆ Cost savings
- ◆ Cost avoidance
- ◆ Deletion of duplicates in inventory
- ◆ Detection of purchase price anomalies
- ◆ Vendor evaluation
- ◆ Identification of functionally equivalent products

Business Case



- ◆ Ontologically-formalized standards are easier to adopt
- ◆ They provide numerous migration, integration, and interoperability advantages
- ◆ An ontological approach identifies knowledge gaps
- ◆ Business communities will enjoy standards that are more stable and provide more rapid returns on investment