Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE
   AUG 2004

2. REPORT TYPE

3. DATES COVERED
   00-00-2004 to 00-00-2004

4. TITLE AND SUBTITLE
   CODIP Technology at Work

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
   AT&T Government Solutions Inc, 1900 Gallows Rd, Vienna, VA, 22182

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
   Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
   Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop (PerMIS ’04), Gaithersburg, MD on August 24-26 2004

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:
   a. REPORT
      unclassified
   b. ABSTRACT
      unclassified
   c. THIS PAGE
      unclassified

17. LIMITATION OF ABSTRACT
   Same as Report (SAR)

18. NUMBER OF PAGES
   10

19a. NAME OF RESPONSIBLE PERSON

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18
The CODIP program provides frameworks and components for intelligent processing of information based on its semantics.

- Distribution of information from publishers to subscribers using subscriber defined semantic queries.
- Automatic generation of semantic mapping between ontologies to facilitate database integration, content translation and distribution.
- Application of a UML technology to leverage existing resources to provide knowledge engineering capability.
- Ontological processing components and services that can bring built-in knowledge processing capability to applications.
Overview
Applications and Products

• Primary products support these applications:

  – **Duet** to support visualization, application and management of ontologies using the UML/MOF engineering standards,
  – **Kage** to support applications with analysis, translation, and repository functionality,
  – **ODKD** for semantics based publication of information to subscribers, and
  – **Artic** to support using multiple ontologies concurrently by finding and codifying relationships between their concepts.

• These products are built from library of reusable components that may be integrated into other applications.

http://codip.grci.com
• **Knowledge Access Engine**
  – Provides management and access to collections of ontologies.

• **Ontology Mapping Engine**
  – Provides automated analysis of potential mappings between ontologies and builds *articulation ontologies* that codify the mappings.

• **Artic Service**
  – Provides APIs, command line and web based access to the mapping engine.
Artic Ontology Mapping Engine

- **Multiple Layers**
  - Non-procedural rules and procedural processors, invoked by rules.
- **Multiple Phases**
  - Analysis, Match Factors, Matches.
Articulation Ontologies

• **Articulations are specialized ontologies that relate concepts in other ontologies.**
  
  – Relationships of various types:
    » Similarity, Part-Of, Kind-Of, Temporal, Spatial, and Domain Specific.
  
  – Multiplicity may be 1:1, 1:M, M:1 or M:M
  
  – Variable ‘strength’.
  
  – May include conversion rules, which may be one-way.

Conversion Rule:
subject.Area := Source.Width *Source.Length
Example Match
Simple, ‘Perfect’

BoatWidth
file:Shipyard_2.owl#genid308

owl:equivalentTo

BoatWidth
file:Shipyard_3.owl#genid311

<Match
kind”="owl:sameIndividualAs"
rdf:about="artic2659"
strength="10.00" >

<MatchFactor rdf:about="artic1913"
detail="BoatWidth"
kind="EX_NM_MTCH:ele_nm"
strength="7" >

<MatchFactor rdf:about="artic2394"
detail="BoatWidth"
kind="EX_NM_MTCH:ele_access_nm"
strength="6" >

<MatchFactor rdf:about="artic2271"
detail="Boat width in meters rounded up."
kind="PHRASE_MTCH:ele_defn"
strength="10" >

<MatchFactor rdf:about="artic2341"
detail="FLOAT8"
kind="EX_DT_MTCH:ele_data_type"
strength="2" >
Articulation Example
Structural, Imperfect Match

Parent Match

WorkOrder
file:Shipyard_2.owl#genid662

owl:equivalentTo

WorkOrder
file:Shipyard_3.owl#genid735

<Match
kind="owl:sameIndividualAs"
rdf:about="artic3027"
strength="8.60" >

<MatchFactor rdf:about="artic1730"
kind="EX_NM_MTCH:tbl_nm"
detail="WorkOrder"
strength="8"> reasons

<MatchFactor rdf:about="artic3029"
detail="artic3027 artic2999"
kind="TBL_ELE_MTCH"
strength="3"> reasons

<MatchFactor rdf:about="artic1819"
detail="Unique marina work order ID." kind="PHRASE_MTCH:ele_defn"
strength="8"> reasons

<MatchFactor rdf:about="artic1825"
detail="INTEGER" kind="EX_DT_MTCH:ele_data_type"
strength="2"> reasons

OrderID
file:Shipyard_3.owl#genid342

owl:equivalentTo

Child Match

WorkOrderID
file:Shipyard_2.owl#genid308

Structural Match
Matching elements found in matching tables generate reciprocal factors

reasons
**I3Con Processing**

- **Pre Processing** applies transformations to
  - Convert DAML and RDFS to compatible OWL equivalents.
  - Adds XML Base namespace if needed.
- **Post Processing** applies transformations to
  - Convert Articulation to Alignment format.
  - Remove low confidence (< 0.8) and uninteresting matches (e.g. ‘genid’).
  - Removes faulty matches (e.g. rdf:ID="" )
I3Con Results Summary

<table>
<thead>
<tr>
<th>Alignments</th>
<th>Animals</th>
<th>Sports</th>
<th>Comsci</th>
<th>Hotel</th>
<th>Network</th>
<th>P&amp;P</th>
<th>P&amp;P-noi</th>
<th>Russia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1.0006</td>
<td>0.9671</td>
<td>0.0000</td>
<td>0.9396</td>
<td>1.0011</td>
<td>1.0012</td>
<td>0.9824</td>
<td>0.9997</td>
</tr>
<tr>
<td>Minimum*</td>
<td>0.8423</td>
<td>0.8423</td>
<td>0.0000</td>
<td>0.8435</td>
<td>0.8423</td>
<td>0.8423</td>
<td>0.8423</td>
<td>0.8423</td>
</tr>
<tr>
<td>Average</td>
<td>0.9283</td>
<td>0.9223</td>
<td>N/A</td>
<td>0.8808</td>
<td>0.9051</td>
<td>0.9111</td>
<td>0.9255</td>
<td>0.9024</td>
</tr>
<tr>
<td>Median</td>
<td>0.9326</td>
<td>0.9384</td>
<td>N/A</td>
<td>0.8709</td>
<td>0.9374</td>
<td>0.9374</td>
<td>0.9358</td>
<td>0.9359</td>
</tr>
</tbody>
</table>

* Confidence < 0.8 removed.

- **Issues**
  - Namespaces – XML Base needed to allow local file usage.
  - ID verses rdf:ID – leads to resources with no ID.
- **Semantic differences between RDF/S, DAML+OIL, OWL**
- **Some results not understood**
  - Comsci topic lead to no alignments
  - Removal of instance data in People&Pets produced more alignments.