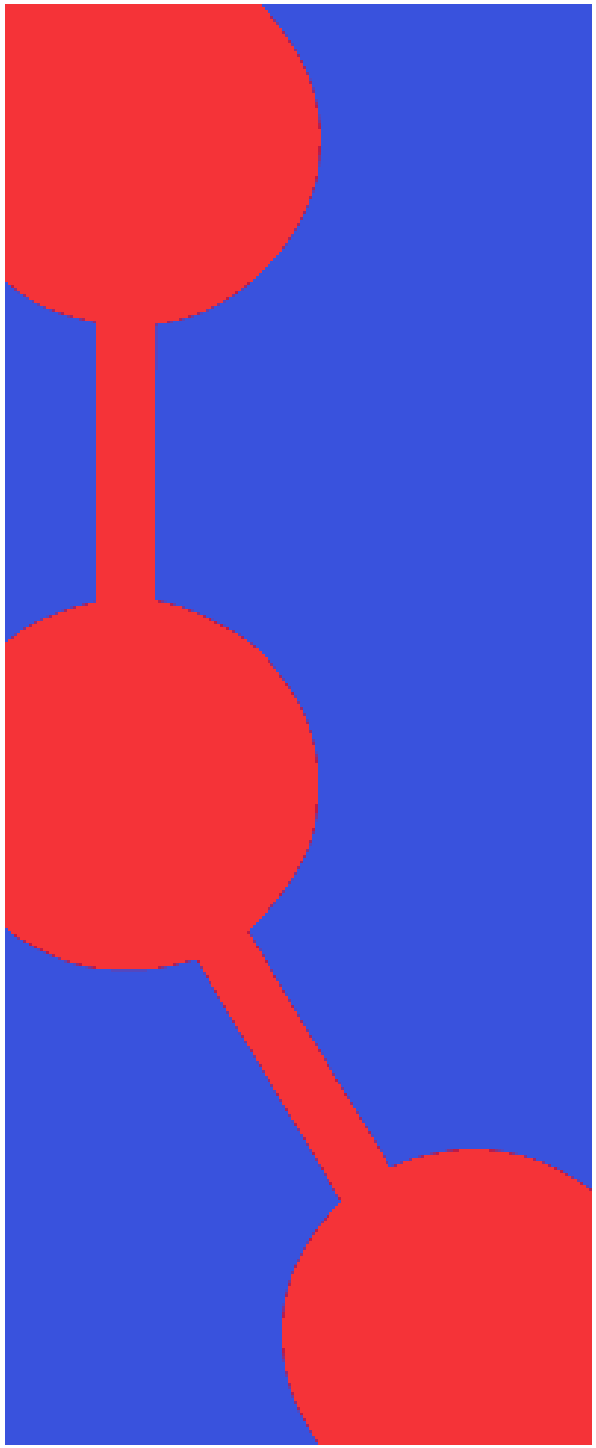
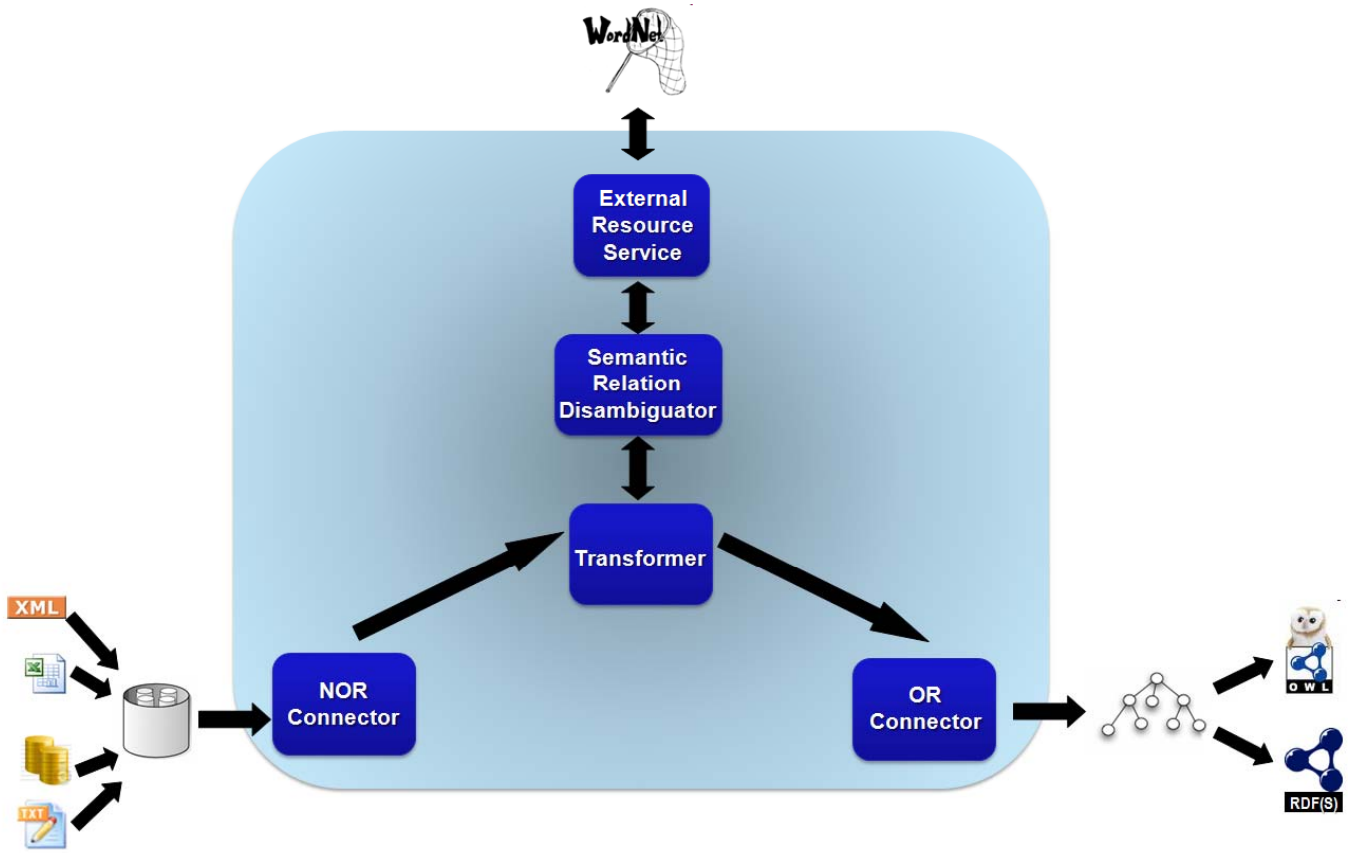


NOR_2O: a Library for Transforming Non-Ontological Resources to Ontologies

Boris Villazón-Terrazas, and **Asunción Gómez Pérez**
Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid
http://www.oeg-upm.net
Phone: 34.91.3366605, Fax: 34.91.3524819

Introduction

- NOR_2O is a software library for transforming non-ontological resources to ontologies.



http://mccarthy.dia.fi.upm.es/nor2o/

For using the External Resource Service you should install WordNet (it is included in the NOR_2O distribution, please read the readme.txt file)

NOR data models

- Non-Ontological Resources
  - Classification Schemes
    - Path Enumeration
    - Adjacency List
    - Snowflake
    - Flattened
  - Thesauri
    - Record based
    - Relation based
  - Lexicons
    - Record based
    - Relation based

How to use the software library?

- There are 3 configuartion files:
  - nor.xml, for describing the non-ontological resource.
  - prnor.xml, for describing the pattern for re-engineering.
  - or.xml, for describing the ontology generated.
- There is a batch file
  - nor2o.bat, for performing the transformation
- Requirements
  - Java runtime environment (tested in 1.5)
  - You should set tha JAVA_HOME

Example: Transforming the ASFA thesaurus – TBox Transformation

- nor.xml, description of the non-ontological resource

```
<Nor type="Thesaurus" name="ASFA">
  <Schema>
    <SchemaEntities>
      <SchemaEntity name="Term">
        <Attribute name="Identifier" valueFrom="DESCRIPTOR" type="string"/>
        <Relation name="NT" using="RecordBased" valueId="NT" destination="Term"/>
        <Relation name="BT" using="RecordBased" valueId="BT" destination="Term"/>
        <Relation name="RT" using="RecordBased" valueId="RT" destination="Term"/>
        <Relation name="UF" using="RecordBased" valueId="UF" destination="NonPreferredTerm"/>
      </SchemaEntity>
      <SchemaEntity name="NonPreferredTerm">
        <Attribute name="Identifier" valueFrom="NON-DESCRIPTOR" type="string"/>
      </SchemaEntity>
    </SchemaEntities>
  </Schema>
  <DataModel>
    <RecordBased>
      <Entity>TERM</Entity>
    </RecordBased>
  </DataModel>
  <Implementation>
    <Xml xmlFile="nors/thesauri/recordbased/asfa_xml_20060522.xml" xsdFile="nors/thesauri/recordbased/asfa.xsd"/>
  </Implementation>
</Nor>
```

Example: Transforming the ASFA thesaurus – TBox Transformation

- prnor.xml, description of the pattern
  - Transformation approach: TBox using WordNet for the semantics of the relations
    - Term -> Class

```
<Prnor identifier="PR-NOR-CLLO-10" transformationApproach="TBox" externalResource="WordNet">
  <Class from="Term" identifier="[Identifier]">
    <ObjectProperty from="RT" to="relatedTerm"/>
    <ObjectProperty from="BT" to="subClassOf"/>
    <ObjectProperty from="UF" to="rdfs:label"/>
  </Class>
</Prnor>
```

- or.xml, description of the ontology generated

```
<Or name="asfa ontology" ontologyURI="http://droz.dia.fi.upm.es/ontologies/asfa.owl"
  ontologyFile="asfa.owl" implementation="OWL" alreadyExist="no">
</Or>
```

Example 2: Industry Production Index, Spreadsheet

Industry Production Index

Year

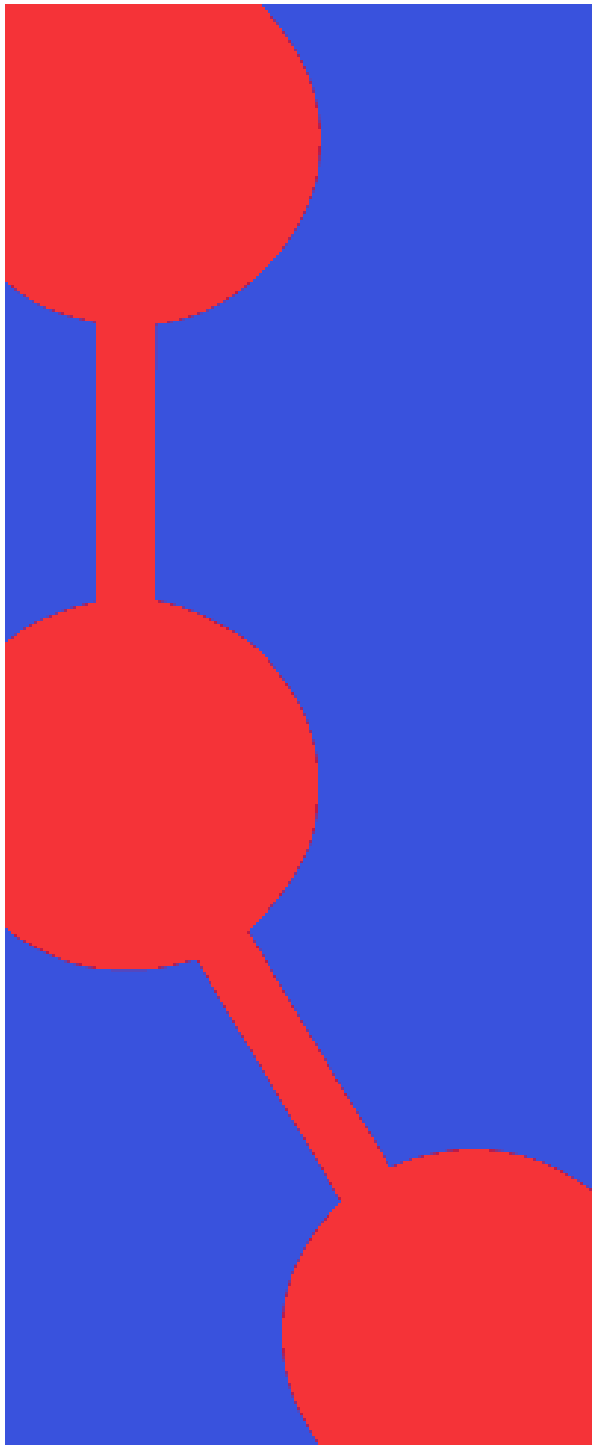
Province

	2009	2008	2007	2006	2005	2004	2003
Total Nacional	3355830	3422239	3336657	3174393	3064129	2942583	2819159
Álava	21988	22316	20870	[illegible]	19638	19779	19638
Albacete	27380	27647	27068	25531	24385	23550	22547
Alicante	136239	142307	140145	133016	123333	113852	111805
Almería	43501	45[illegible]0	43970	40871	38766	36260	33947
Asturias	71853	73124	72276	70115	68175	67088	65062
Ávila	11455	11708	11434	10900	10611	10319	10211
Badajoz	40874	41358	40168	38045	37052	34972	34866
Illes Balears	91824	93335	91254	88027	87024	85425	75951
Barcelona	467335	477942	469432	444410	436294	417425	397693
Burgos	25567	25891	25372	24504	23733	22882	22159
Cáceres	26307	26494	26064	25039	24846	20596	23[illegible]40
Cádiz	62817	64505	63338	61691	58986	57138	54[illegible]2
Cantabria	39611	40393	39560	37690	36561	35649	34017
Castellón	42122	43855	42476	39749	37865	37214	3421[illegible]
Ciudad Real	32046	33011	31881	30446	29521	29011	26778
Córdoba	48979	50057	49302	47155	45405	43394	41964
Coruña, A	83748	84220	82873	79170	77023	74809	71748
Cuenca	14747	14928	14741	13822	13336	12829	12546
Girona	58404	51467	50108	47169	46827	45145	52482
Granada	60016	62269	61055	57223	54341	50508	4966[illegible]
Guadalajara	13507	13735	12874	11825	10438	10120	942[illegible]
Guipúzcoa	52034	63569	59546	58486	57193	56498	55913
Huelva	26783	27463	27063	25487	24777	24270	22947
Huesca	17837	17109	16694	16025	15390	15078	14283
Jaén	36157	37368	36962	35383	34675	33157	32444
León	33564	34012	33563	32359	31664	30992	30256
Lleida	36920	37638	36065	33956	32739	31515	29605
Lugo	24861	25035	24609	23780	23122	22479	22396
Madrid	511804	519307	503000	478202	456175	436074	407655
Málaga	113362	116583	114547	108713	102382	96087	88257
Murcia	95636	100074	97374	90698	85110	82484	75973
Navarra	43282	43847	43142	41083	40739	39679	38936
Ourense	23304	23711	23526	22843	22452	22118	21560
Palencia	10964	11111	11060	10694	10575	10399	10297

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- <nor xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLo
<Nor type="GenericSpreadSheet" name="Spanish Employment">
  <Schema>
    <SchemaEntities>
      <SchemaEntity name="Location">
        <Attribute name="Name" valueFrom="pcaxis.[A11:A62]" type="string"/>
      </SchemaEntity>
      <SchemaEntity name="Dataset">
        <Attribute name="Name" valueFrom="pcaxis.[B6:H6]" type="string"/>
      </SchemaEntity>
      <SchemaEntity name="Year">
        <Attribute name="Name" valueFrom="pcaxis.[B9:H9]" type="string"/>
      </SchemaEntity>
      <SchemaEntity name="IndustryProductionIndex" type="Nary">
        <Attribute name="hasValue" valueFrom="pcaxis.[B10:H62]" type="string"/>
        <Relation name="inArea" usingSpreadSheetColumn="A" destination="Location"/>
        <Relation name="inPeriod" usingSpreadSheetRow="9" destination="Year"/>
        <Relation name="dataset" usingSpreadSheetRow="6" destination="Dataset"/>
      </SchemaEntity>
    </SchemaEntities>
  </Schema>
  <DataModel>
    <GenericDataModel/>
  </DataModel>
  <Implementation>
    <Spreadsheet type="ms" file="Empresas.xls"/>
  </Implementation>
</Nor>
```

NOR$_2$O

```
<!-- http://geo.linkeddata.es/resource/Estad%C3%ADstica/Madrid_ipi_2003 -->
<scv:Item rdf:about="Estad%C3%ADstica/Madrid_ipi_2003">
  <rdfs:label xml:lang="es">&#205;ndice de Producci&#243;n Industrial de Madrid en el a&#241;o 2003</rdfs:label>
  <rdf:value rdf:datatype="&xsd;double">407655.0</rdf:value>
  <scv:dimension rdf:resource="A%C3%B1o/2003"/>
  <scv:dataset rdf:resource="Estad%C3%ADstica/%25C3%258DndiceDeProducci%25C3%25B3nIndustrial"/>
  <onto:inProvincia rdf:resource="Provincia/Madrid"/>
</scv:Item>
<!-- http://geo.linkeddata.es/resource/Estad%C3%ADstica/Madrid_ipi_2004 -->
<scv:Item rdf:about="Estad%C3%ADstica/Madrid_ipi_2004">
  <rdfs:label xml:lang="es">&#205;ndice de Producci&#243;n Industrial de Madrid en el a&#241;o 2004</rdfs:label>
  <rdf:value rdf:datatype="&xsd;double">436074.0</rdf:value>
  <scv:dimension rdf:resource="A%C3%B1o/2004"/>
  <scv:dataset rdf:resource="Estad%C3%ADstica/%25C3%258DndiceDeProducci%25C3%25B3nIndustrial"/>
  <onto:inProvincia rdf:resource="Provincia/Madrid"/>
</scv:Item>
```



NOR_2O: a Library for Transforming Non-Ontological Resources to Ontologies

Boris Villazón-Terrazas, and **Asunción Gómez Pérez**

Facultad de Informática, Universidad Politécnica de Madrid

Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

http://www.oeg-upm.net

bvillazon at fi.upm.es

Phone: 34.91.3366605, Fax: 34.91.3524819