



Transitioning Applications to Ontologies

*Kalina Bontcheva
on behalf of
the TAO consortium*

<http://www.tao-project.eu>

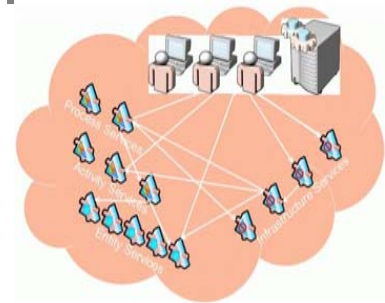
Funded by: European Commission – 6th Framework
Project Reference: IST-2004-026460



- Addressing the problem of **transitioning legacy applications to ontologies**
- What is a legacy software system:
 - “A large software system that is vital to [an] organisation, but resists modification and evolution to meet new and constantly changing business requirements”
- Towards **semantic-assisted software engineering**

Legacy systems: Main Problems

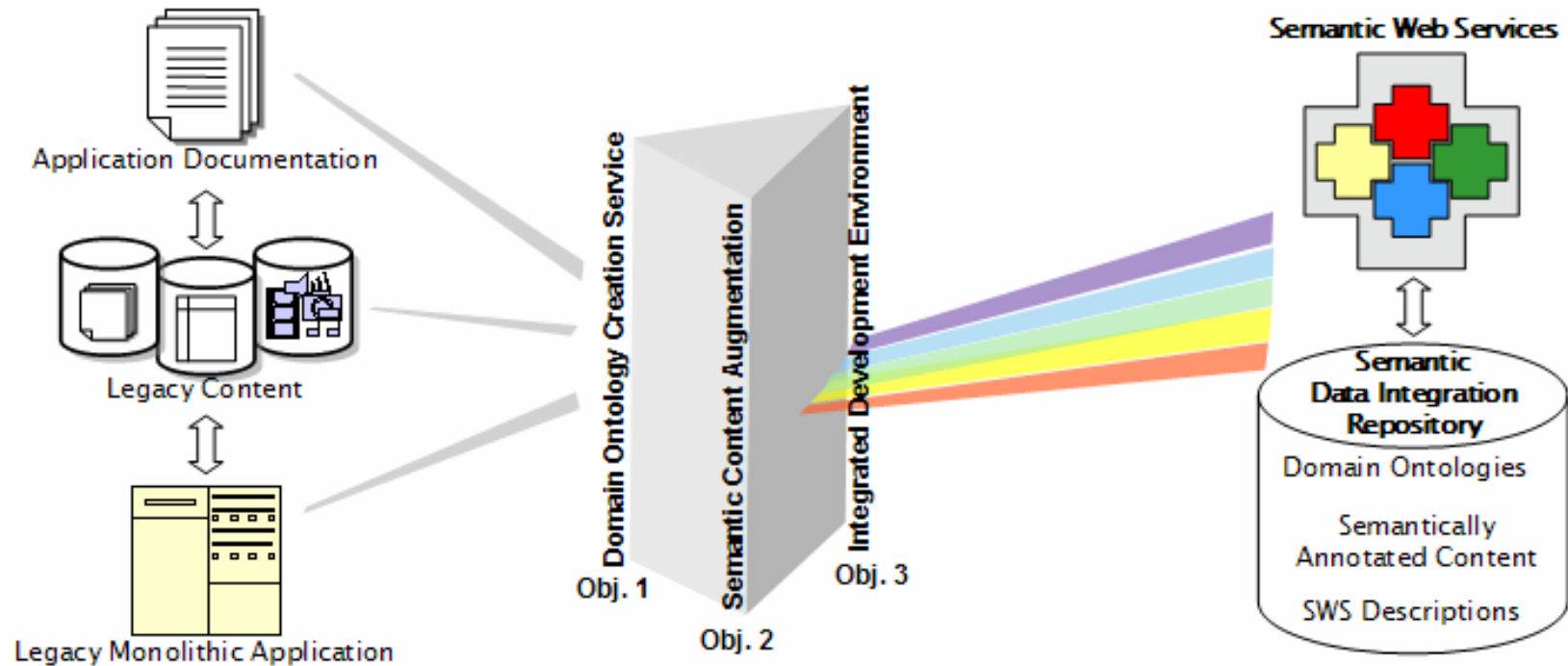
- Enterprise Application Integration:
 - Built with languages and data models that are now **out-dated**
 - Badly structured and **hard to maintain**
 - Badly documented and understood
 - Difficult to integrate with each other and with new systems
 - Need for migration towards Web 2.0 applications & services



TAO: Towards a Low-Cost Migration Path

- Make transitioning to ontologies **fast and effective**
- Build a **reusable transitioning process**
- **Minimize** consulting **time** during migration and integration
- **Minimize costs**
- **Reduce integration** overheads and **limit risk**

Transitioning Process



- **Semi-automatic learning of domain ontologies** from software artefacts and legacy content
- **Semantic augmentation of legacy content** and web service definitions
- **Heterogeneous & distributed** semantic repositories
- **Transitioning Infrastructure supporting IT providers**

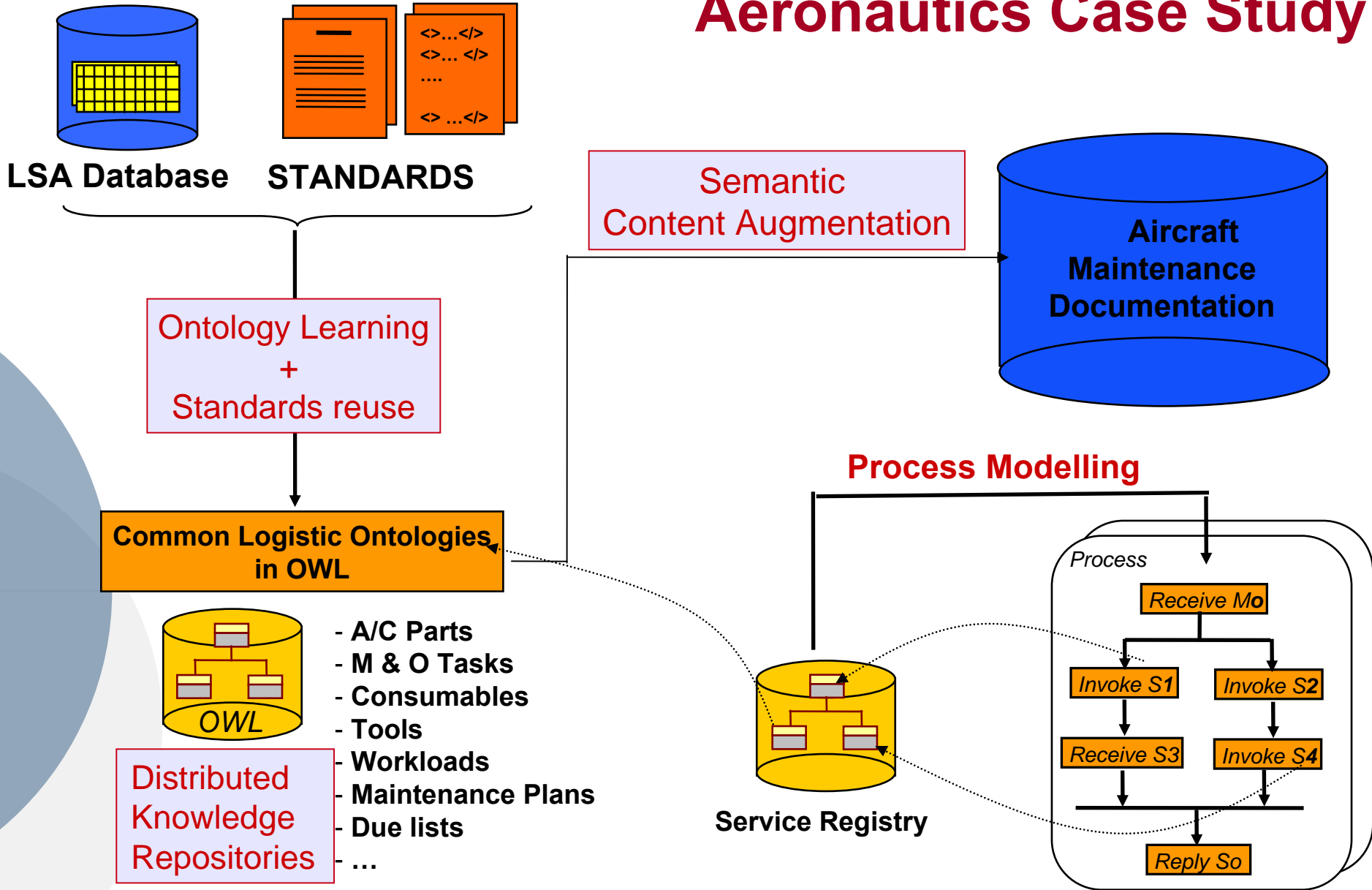
Partner Contributions

-  The University of Sheffield. Semantic annotation of text and multimedia content; GATE – widely used open-source infrastructure for text mining
-  University of Southampton Transitioning methodology
-  Institut "Jožef Stefan" Machine learning and data mining
-  Atos Origin Transitioning infrastructure; exploitation
-  Ontotext Semantic Technology Lab Most scalable & efficient semantic repository
-  MONDECA Content and knowledge management
-  DASSAULT AVIATION Aviation case study; aircraft manufacturer

Transitioning Problems

- **From legacy databases to ontologies**
- ◆ Towards semantic-based software engineering
- ◆ Transitioning web applications towards ontologies and services

Aeronautics Case Study



Improving *semantic interoperability* in business processes through *Semantic technologies*

Example Learnt Ontology – 400k triples

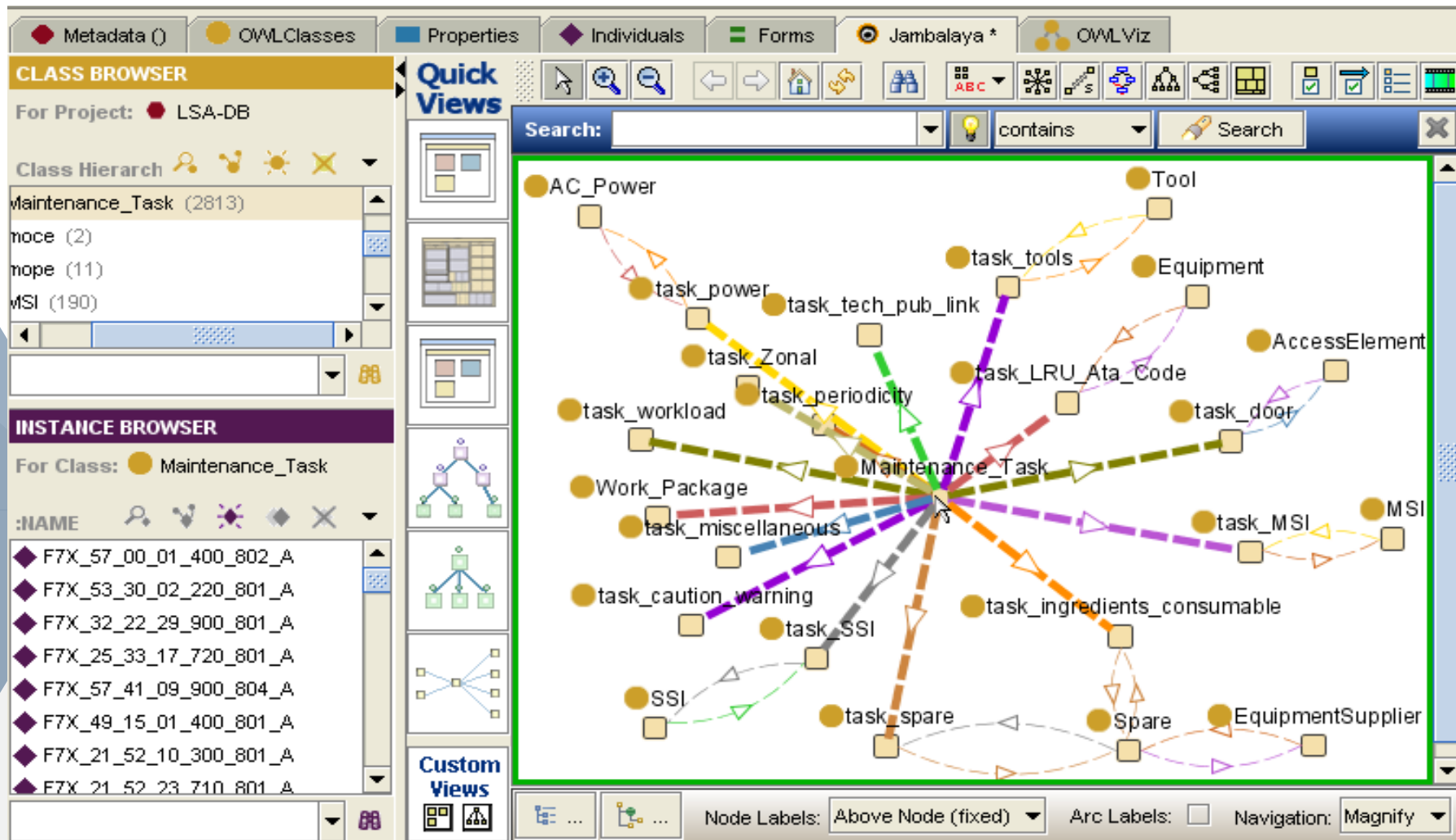
The screenshot displays a web-based ontology editor interface with two main panels: **SUBCLASS EXPLORER** and **CLASS EDITOR**.

SUBCLASS EXPLORER: Shows the project name "LSA-DB" and an "Asserted Hierarchy" tree. The tree lists various classes, with "Tool" expanded to show its subclasses: "Covers_plugs_blanking_plates", "Cranes_Supports", "Electrical_engineering", "Electronics_electrical_engineering", "Fixed_Size_and_adjustable_wrenches", "General_maintenance", and "GSE".

CLASS EDITOR: Shows the selected class "Covers_plugs_blanking_plates" (instance of owl:Class). It includes a table for properties and values, currently showing "rdfs:comment". Below this, a list of subclasses is displayed, each with a checkbox and a data type: "ATO_Aircraft (single string)", "ATO_ATA_num (single string)", "ATO_Condition_ID (single string)", "ATO_description (single string)", "ATO_Falcon (single string)", "ATO_Logistic_Number (single string)", "ATO_nato_vendor_code (single string)", "ATO_NUM (single int)", and "ATO_on_board (single string)".

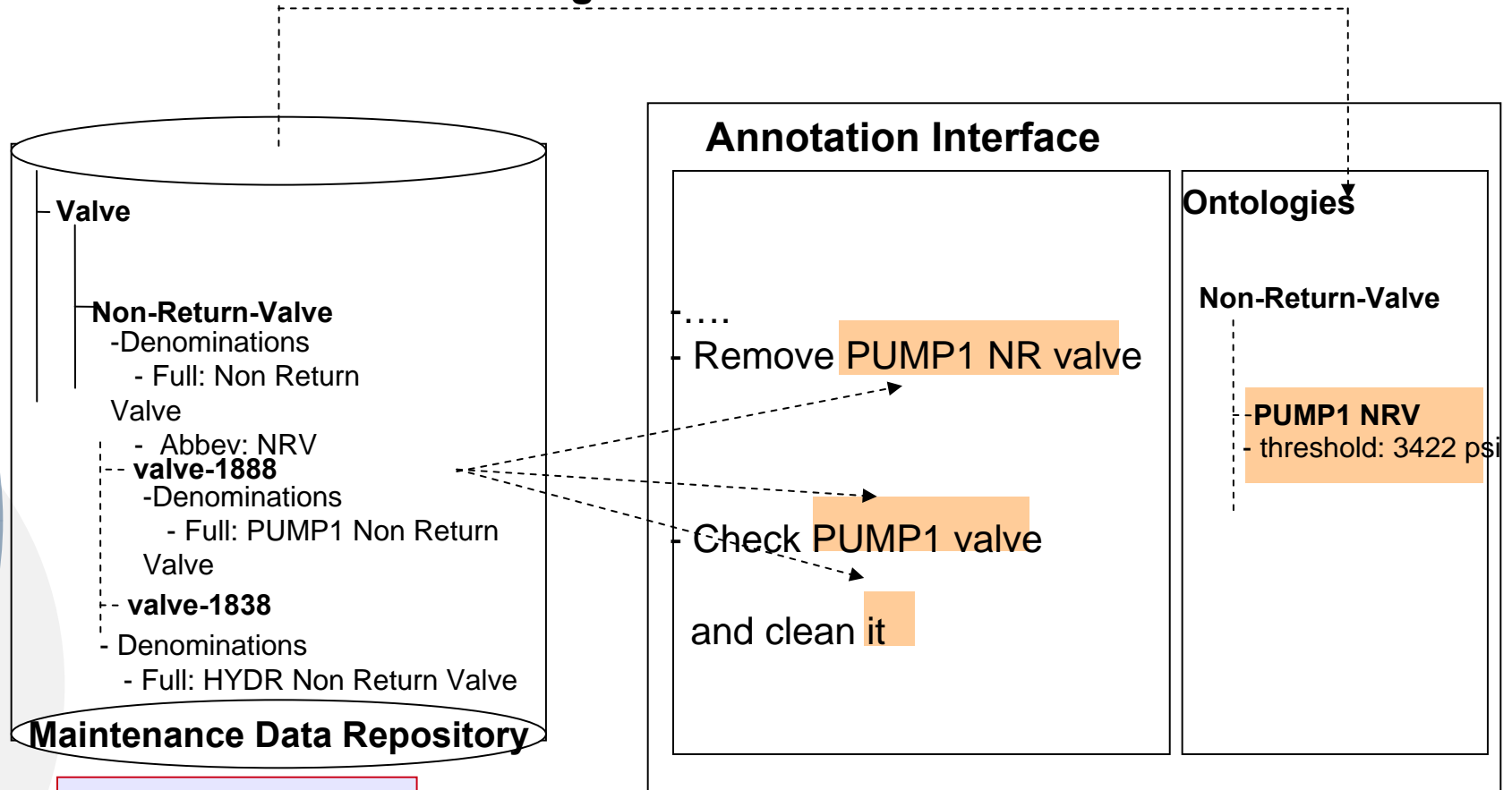
At the bottom right, there is a "Superclasses" section showing "Tool" as a superclass. The interface also features a bottom toolbar with icons for navigation and a view selector for "Logic View" and "Properties View".

Example Learnt Ontology (2)



A Semantic Tagging Example

Content Augmentation of Maintenance Manuals



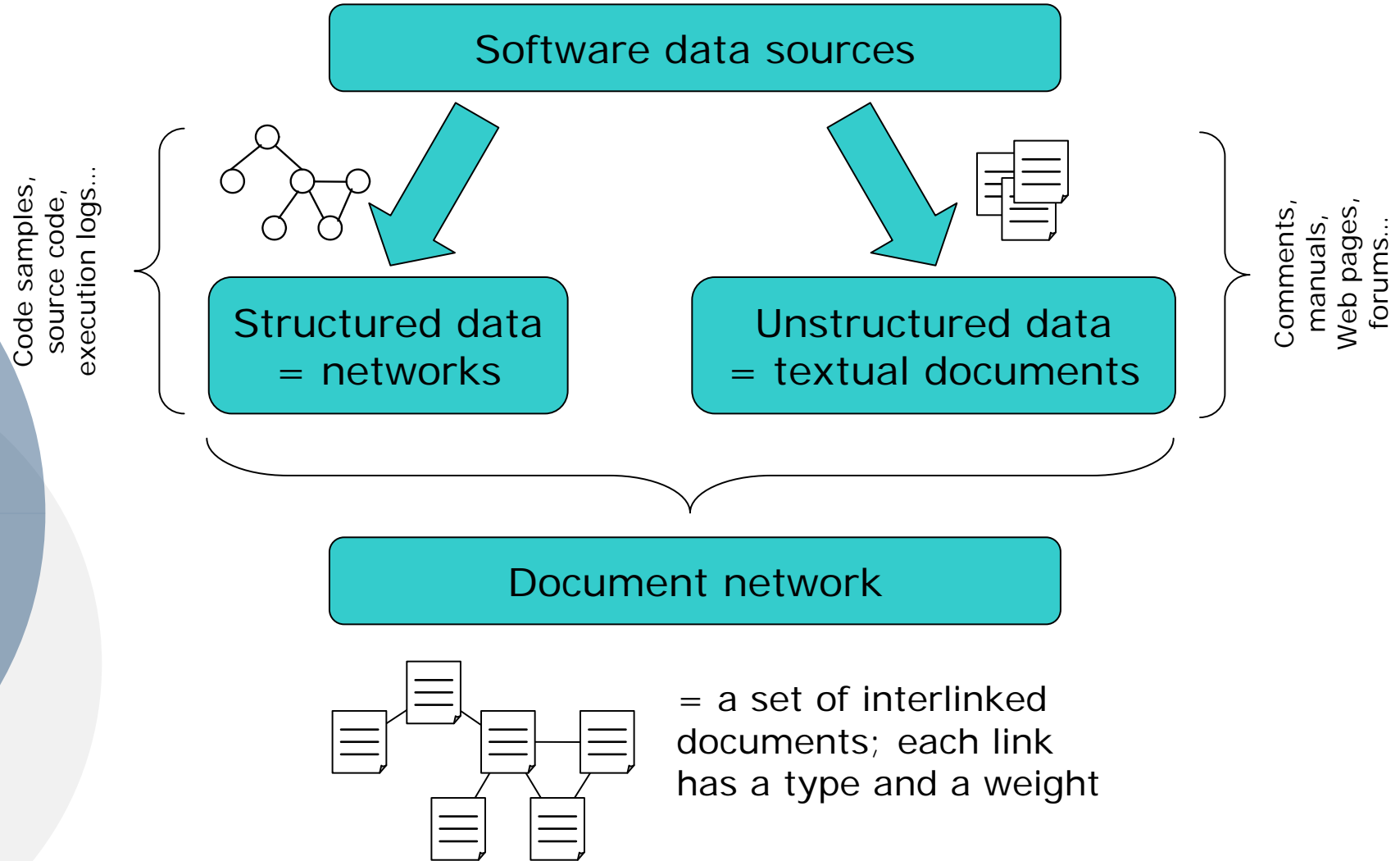
Heterogeneous Knowledge Store

Semantic Content Augmentation

Transitioning Problems

- ✓ From legacy databases to ontologies
- Towards semantic-based software engineering
- ◆ Transitioning web applications towards ontologies and services

Learning Ontologies from Software Artefacts



Creating a Document Network

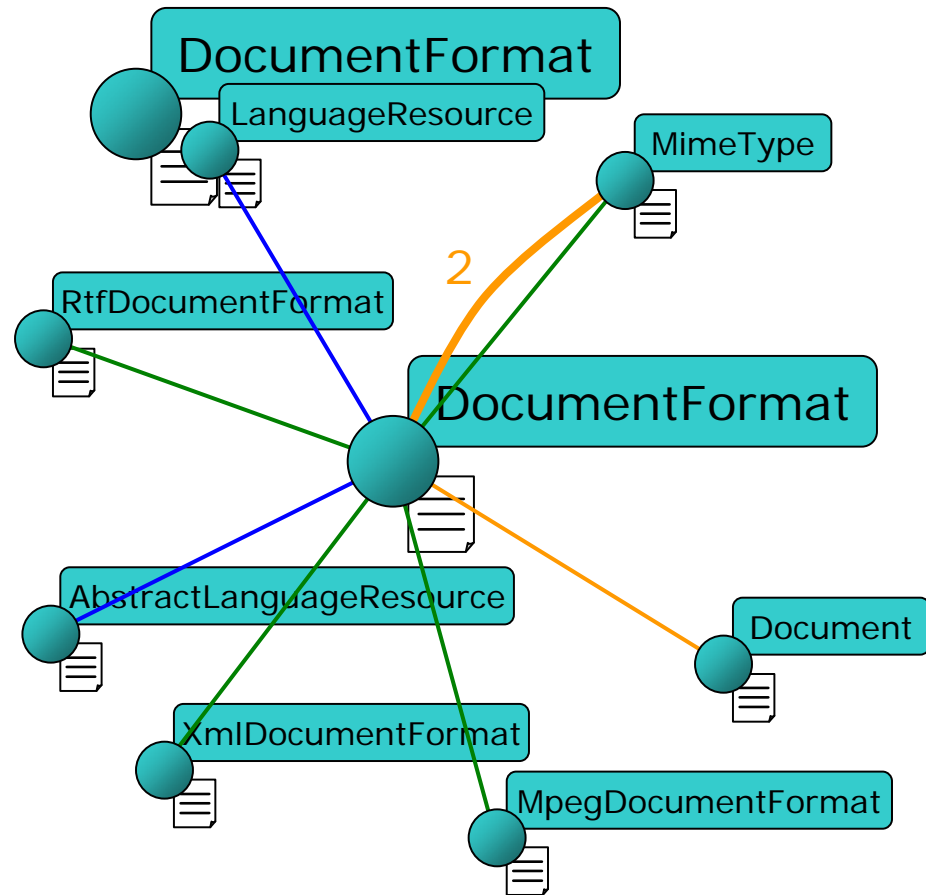
DocumentFormat.class

```
/** The format of Documents. Subclasses of DocumentFormat know about
 * particular MIME types and how to unpack the information in any
 * markup or formatting they contain into GATE annotations. Each MIME
 * type has its own subclass of DocumentFormat, e.g. XmlDocumentFormat,
 * RtfDocumentFormat, MpegDocumentFormat. These classes register themselves
 * with a static index residing here when they are constructed. Static
 * getDocumentFormat methods can then be used to get the appropriate
 * format class for a particular document.
 */
public abstract class DocumentFormat
extends AbstractLanguageResource implements LanguageResource(

/** The MIME type of this format. */
private MimeType mimeType = null;

/**
 * Find a DocumentFormat implementation that deals with a particular
 * MIME type, given that type.
 * @param aGateDocument this document will receive as a feature
 * the associated Mime Type. The name of the feature is
 * MimeType and its value is in the format type/subtype
 * @param mimeType the mime type that is given as input
 */
static public DocumentFormat getDocumentFormat(gate.Document aGateDocument,
MimeType mimeType){

} // getDocumentFormat(aGateDocument, MimeType)
} // class DocumentFormat
```



Ontology Learnt from Software Code

Messages OWLIM Ontology LR_00016

Classes & Instances Properties

Classes and Instances

- DD
- HC
- IR
- KB
- NA
- VT
- GATEPlugin**
 - ANNIE**
 - Chemistry_Tagger
 - Information_Retrieval
 - Jape_Compiler
 - Kea
 - Machine_Learning
 - Minipar
 - Minorthird
 - Montreal_Transducer
 - NLG_Tools
 - NP_Chunking
 - Ontology_Tools
 - SUPPLE
 - Stemmer
 - Tools
 - TreeTagger
 - WordNet
 - annic
 - arabic
 - cebuano
 - crawl
 - creole

Property Types

- isDefinedBy [ALL RESOURCES]
- comment [ALL RESOURCES]
- label [ALL RESOURCES]
- containsResource [GATEResource]
- transitiveOver [ALL CLASSES]
- seeAlso [ALL RESOURCES]
- title [ALL CLASSES]
- versionInfo [ALL RESOURCES]

Property Values

label	ANNIE	X
containsResource	ANNIEAnnotationSchema	X
containsResource	ANNIEGATEUnicodeTokeniser	X
containsResource	ANNIEANNIEEnglishTokeniser	X
containsResource	ANNIEANNIEGazetteer	X
containsResource	ANNIEHashGazetteer	X
containsResource	ANNIEJapeTransducer	X
containsResource	ANNIEANNIENETransducer	X
containsResource	ANNIEANNIESentenceSplitter	X
containsResource	ANNIEANNIEPOSTagger	X
containsResource	ANNIEANNIEOrthoMatcher	X
containsResource	ANNIEANNIEPronominalCoreferencer	X
containsResource	ANNIEANNIENominalCoreferencer	X
containsResource	ANNIEDocumentResetPR	X
containsResource	ANNIEJapeViewer	X
containsResource	ANNIEGaze	X

GATE Ontology Editor Initialisation Parameters

Google-like Conceptual Search (1)

marks Tools Help

http://localhost:8080/cloneservice/search.jsp

Google

The Postgr... Scopus - Marketing H... Researchers BT Home Hub - Home Sustrans Get cycling Sheffield City Council ... GIS Training YSO - Semantic Comp...

Search knowledge with CLOnE QL

Ontology:

Query:

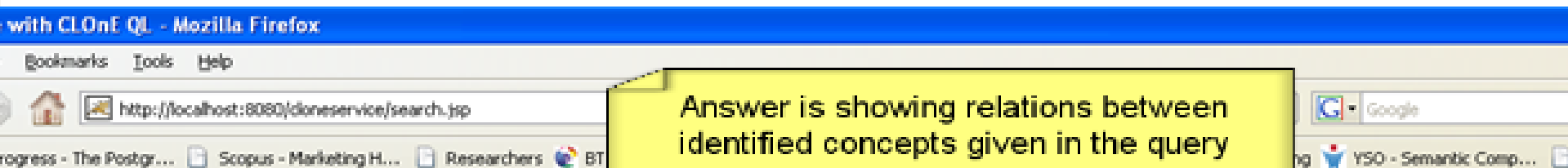
Ask

```
ANNIE POS Tagger --> hasRunTimeParameter --> document
ANNIE POS Tagger --> hasRunTimeParameter --> inputASName
ANNIE POS Tagger --> hasRunTimeParameter --> ANNIEANNIEPOSTaggerOutputASName
ANNIE POS Tagger --> hasRunTimeParameter --> baseTokenAnnotationType
ANNIE POS Tagger --> hasRunTimeParameter --> baseSentenceAnnotationType
ANNIE POS Tagger --> hasRunTimeParameter --> ANNIEANNIEPOSTaggerOutputAnnotationType

Minipar Wrapper --> hasRunTimeParameter --> MiniparMiniparWrapperDocument
Minipar Wrapper --> hasRunTimeParameter --> miniparDateDir
Minipar Wrapper --> hasRunTimeParameter --> MiniparMiniparWrapperMiniparBinary
Minipar Wrapper --> hasRunTimeParameter --> annotationInputSetName
Minipar Wrapper --> hasRunTimeParameter --> MiniparMiniparWrapperAnnotationOutputSetName
```

Result:

Question-Based Conceptual Search (2)



Search knowledge with CLOnE QL

Ontology:

Query:

Ask

Result:

```
ANNIEANNIEPOSTaggerRulesURL --> [inverseProperty] hasInitTimeParameter --> ANNIE POS Tagger
ANNIEANNIEPOSTaggerEncoding --> [inverseProperty] hasInitTimeParameter --> ANNIE POS Tagger
lexiconURL --> [inverseProperty] hasInitTimeParameter --> ANNIE POS Tagger

baseSentenceAnnotationType --> [inverseProperty] hasRunTimeParameter --> ANNIE POS Tagger
outputAnnotationType --> [inverseProperty] hasRunTimeParameter --> ANNIE POS Tagger
ANNIEANNIEPOSTaggerInputASName --> [inverseProperty] hasRunTimeParameter --> ANNIE POS Tagger
ANNIEANNIEPOSTaggerOutputASName --> [inverseProperty] hasRunTimeParameter --> ANNIE POS Tagger
ANNIEANNIEPOSTaggerBaseTokenAnnotationType --> [inverseProperty] hasRunTimeParameter --> ANNIE POS Tagger
Tagger
ANNIEANNIEPOSTaggerDocument --> [inverseProperty] hasRunTimeParameter --> ANNIE POS Tagger
```

Transitioning Problems

- ✓ From legacy databases to ontologies
- ✓ Towards semantic-based software engineering
- Transitioning web applications towards ontologies and services

Transitioning Web Applications

- Legacy application:
 - database driven
 - no interoperability
- Ontologies + SOA:
 - **Learn ontologies**
 - Manage **complex resources** and **knowledge links**
 - Use Service Oriented Application to **integrate added value services** from other suppliers: cartography, translation, booking services...

LIST OF TOP LECTURES (counting since March 2007):



Machine Learning
Probability and
Models

Sam Roweis
1741 views 4
8 comments

BIENVENUE /UR NOTRE TERRITOIRE

- REBERGEMENT
- RESTAURATION
- ARTISAN/PRODUCTEUR
- ACTIVITES SPORTIVES
- PATRIMOINE CULTUREL
- PATRIMOINE NATUREL
- EVENEMENTS
- PERSONALITES

WS 1

WS 2

WS 3

WS 4

WS 5

Tourism resources & links
Ontology



CRM
BOOKING
...

Software Demonstrations on TAO Web site

- Learning ontologies from software code
- Supporting software developers with conceptual search
- RDB2Onto

Thank you! Questions?

This presentation + demos:

www.tao-project.eu/demos-dec07/

Kalina Bontcheva:

kalina@dcs.shef.ac.uk

Contact Information

For queries / further information, please contact the project co-ordinator:

Kalina Bontcheva

Department of Computer Science
University of Sheffield
Regent Court
211 Portobello Street
Sheffield S1 4DP

phone: (+44 - 114) 222 1930

fax: (+44 - 114) 222 1810

email: K.Bontcheva@dcs.shef.ac.uk