



# Transforming an OWL Ontology to an OpenOffice Document Template

- ▶ **Massoud Toussi, MD, Msc, PhD Candidate**
- ▶ Medical Informatics and Bioinformatics Laboratory (LIM&BIO EA 3969), UFR SMBH, Université Paris 13, Bobigny, France  
Formerly: Physician, Editor-in-Chief of Spectrum Medical Journal



# Topics

- ▶ Writing and retrieving scientific papers
- ▶ The potential advantage of Open Document Format (ODF) for semantic content management
- ▶ Paper Element Nodes (PEN): an ontology of scientific papers in the domain of biomedical science written in Web Ontology Language (OWL) using *Protégé*
- ▶ Transforming PEN to an OpenOffice Template



# Writing Scientific Papers

- ▶ Not a trivial task
- ▶ The most difficult thing is structuring the ideas
- ▶ In biomedical science: IMRaD:
  - ▶ Introduction
  - ▶ Methods
  - ▶ Results
  - ▶ Discussion
- ▶ Typing, editing, spell checking, defining styles, layout, inserting tables, figures, managing references, etc



# Retrieving Scientific Papers

- ▶ Today's scientific search engines are capable of searching inside the documents : Google Scholar, or PubMed or EmBase
- ▶ This search does not take into account the semantic structure of the articles, for example:
  - ▶ When searching for Python language, there are articles about the snake
  - ▶ When searching for Java language, there are articles about coffee, and the province
  - ▶ When searching about pain, there are articles about bread, because pain in French means bread



# Two kinds of solutions

- ▶ Making more sophisticated search engines able to « understand » the documents
  - ▶ Dependency to native languages
  - ▶ The complexity of the solution, false negatives
- ▶ Making documents more « understandable » by integrating the semantic structure of articles into documents
  - ▶ Easy to do
  - ▶ More efficient



# Today's Office Suits

- ▶ Facilitate writing of scientific papers in many ways:
  - ▶ Typing and formatting the text layout
  - ▶ Drawing graphics, and making tables
  - ▶ Spell checking
  - ▶ Tracing revisions
  - ▶ Making bibliographies
- ▶ It will be interesting to make tools for
  - ▶ helping authors organize their ideas
  - ▶ helping search engines find portions of articles using their semantic structure



# The Open Document Format (ODF)

- ▶ Is based on XML format
- ▶ Is an ISO/IEC International Standard
- ▶ Shared by many office suites, including:
  - ▶ OpenOffice.org
  - ▶ Koffice
  - ▶ Google Docs
  - ▶ NeoOffice
  - ▶ Zoho
  - ▶ IBM Lotus Symphony
  - ▶ Corel WordPerfect Office X4





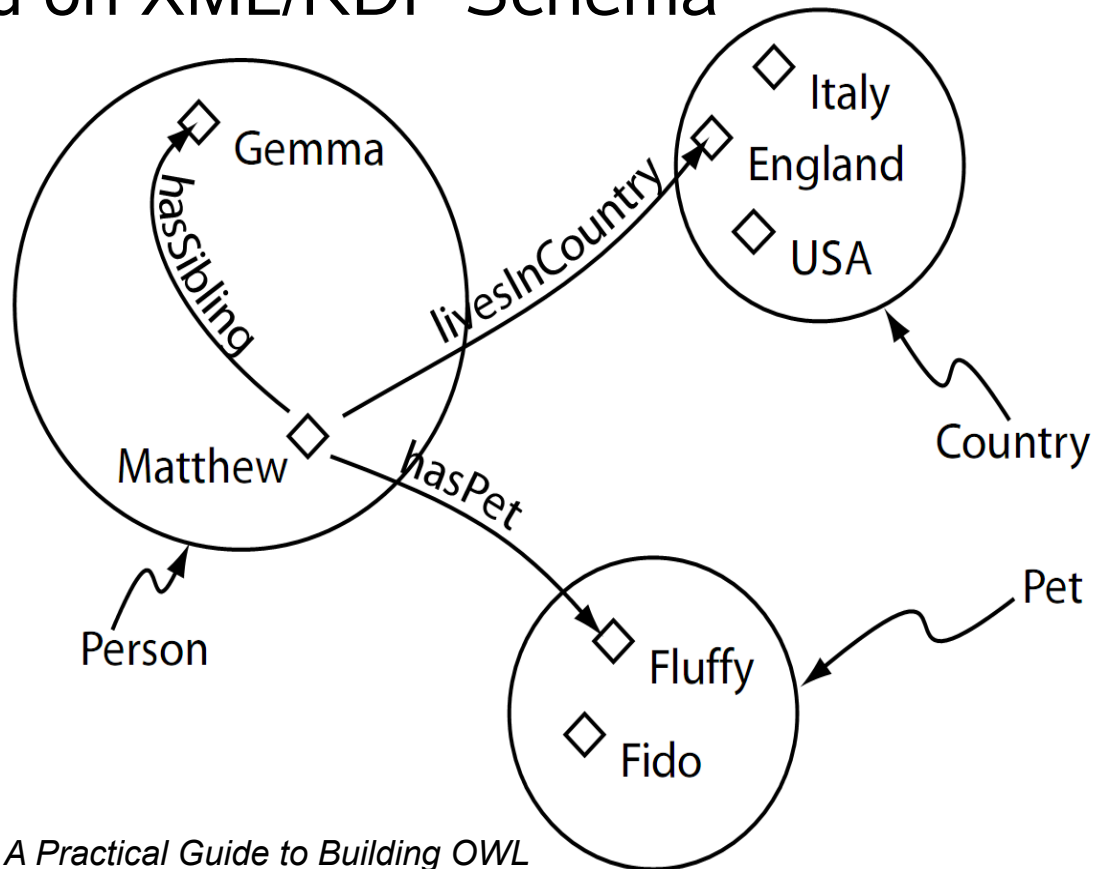
# PEN Ontology

- ▶ PEN (Paper Element Nodes) is an ontology of scientific papers in biomedical domain which describes the way different parts (elements) of an article are grouped in different sections (nodes) by use of OWL language.
- ▶ What is an ontology?
  - ▶ A formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain.



# OWL (Web Ontology Language)

- ▶ A family of knowledge presentation languages for authoring ontologies
- ▶ One of the fundamental technologies of Semantic Web which is based on XML/RDF Schema



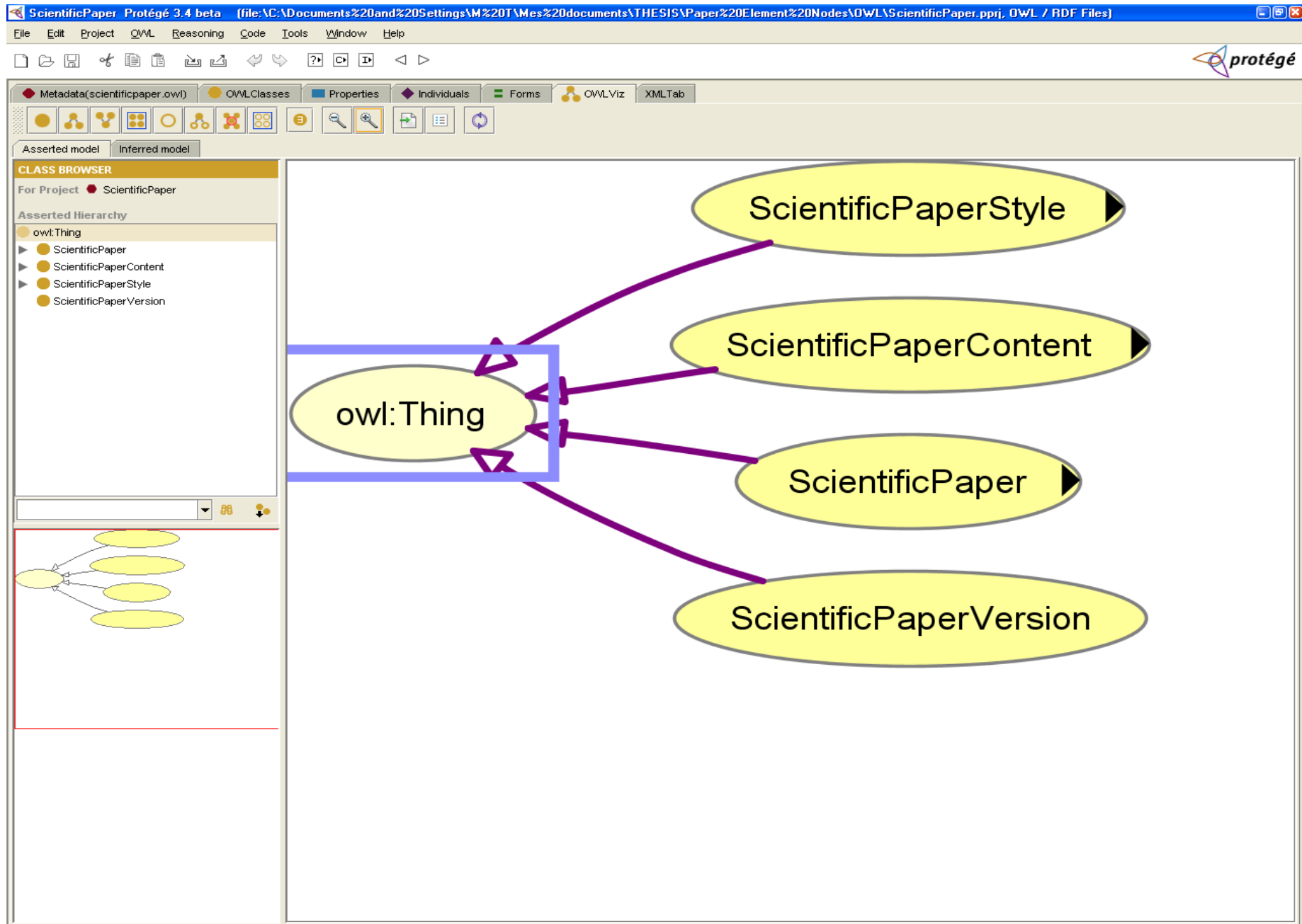
Picture cutesy of: Horridge M et al. *A Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools*, 1.1rd ed, Manchester: University of Manchester; 2007.



# The formation of PEN Ontology

- ▶ Based *Uniform Requirements for Articles Submitted to Biomedical Journals*
- ▶ Completed with other methodological articles to define a **maximum data set**
- ▶ Defining an ontology based on this maximum dataset
  - ▶ Demo: PEN in *Protégé*

# PEN Ontology Main Classes



# PEN Ontology Properties

ScientificPaper Protégé 3.4 beta (file:\C:\Documents%20and%20Settings\M%20T\Mes%20documents\THESES\Paper%20Element%20Nodes\OWL\ScientificPaper...)

File Edit Project OWL Reasoning Code Tools Window Help

Metadata(scientificpaper.owl) OWLClasses Properties Individuals Forms

**PROPERTY BROWSER**  
For Project: ScientificPaper  
Object Datatype Annotation All  
Object properties  
hasContent ↔ isContentOf  
hasStyle ↔ isStyleOf  
hasVersion ↔ isVersionOf  
isContentOf ↔ hasContent  
isStyleOf ↔ hasStyle  
isVersionOf ↔ hasVersion  
Super Properties

**PROPERTY EDITOR for hasContent (instance of owl:Object...)**  
For Property: <http://www.owl-ontologies.com/Ontology1215332743.owl#ha>  
Annotations

Property	Value	Lang
rdfs:comment		

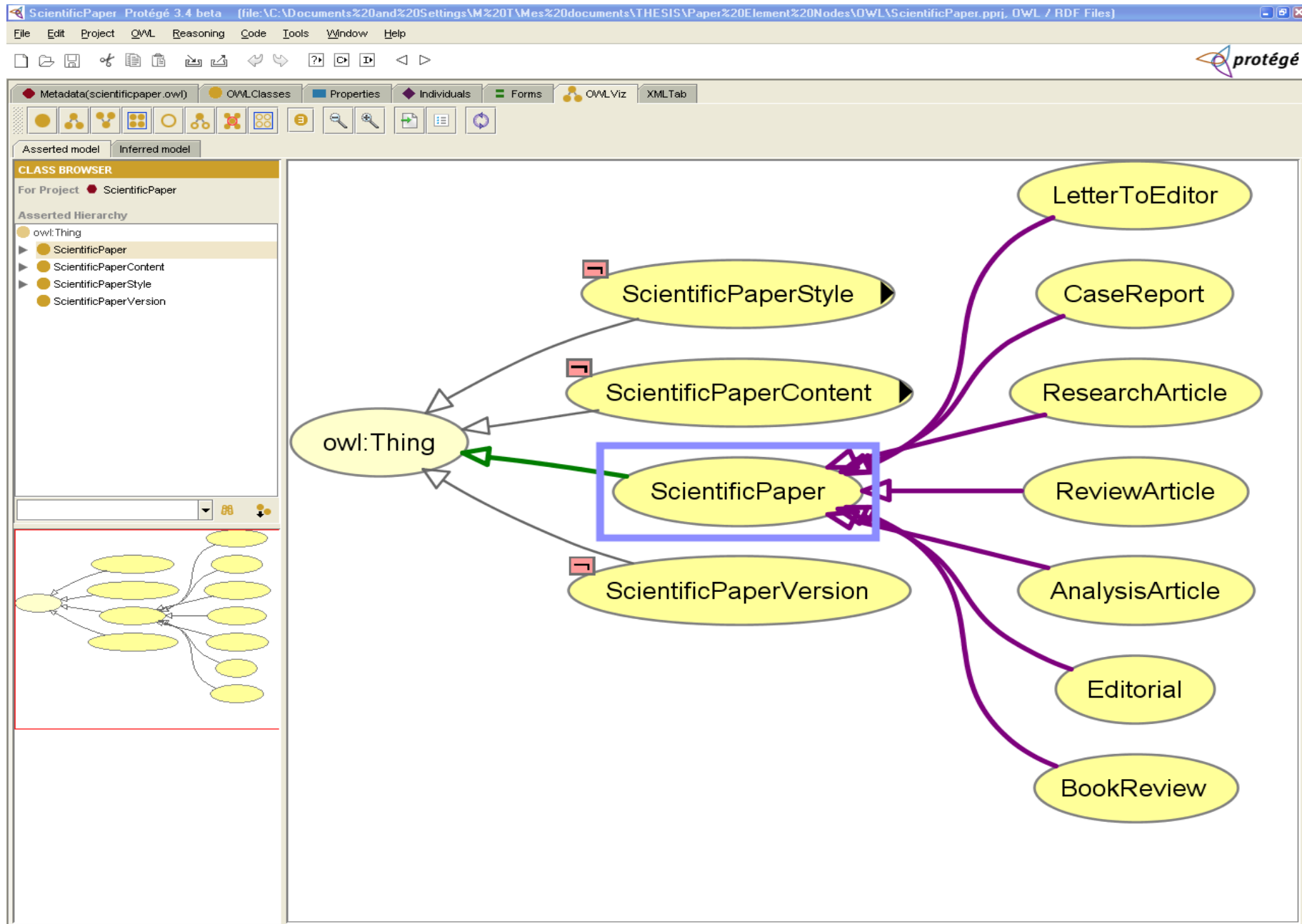
Domain  Range

owl:Thing

Functional  
 InverseFunctional  
 Symmetric  
 Transitive

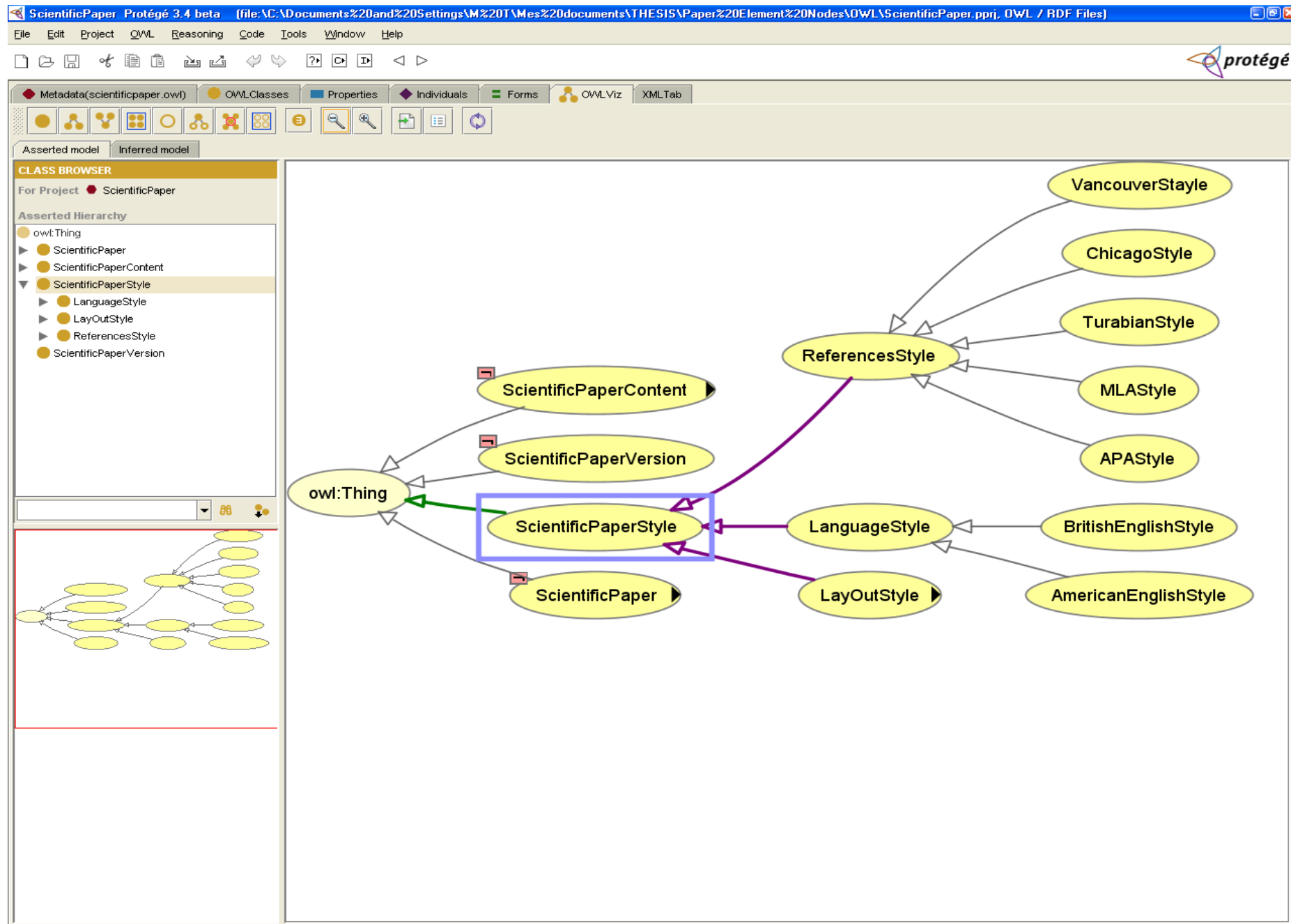
Inverse  isContentOf

# Scientific Paper Class



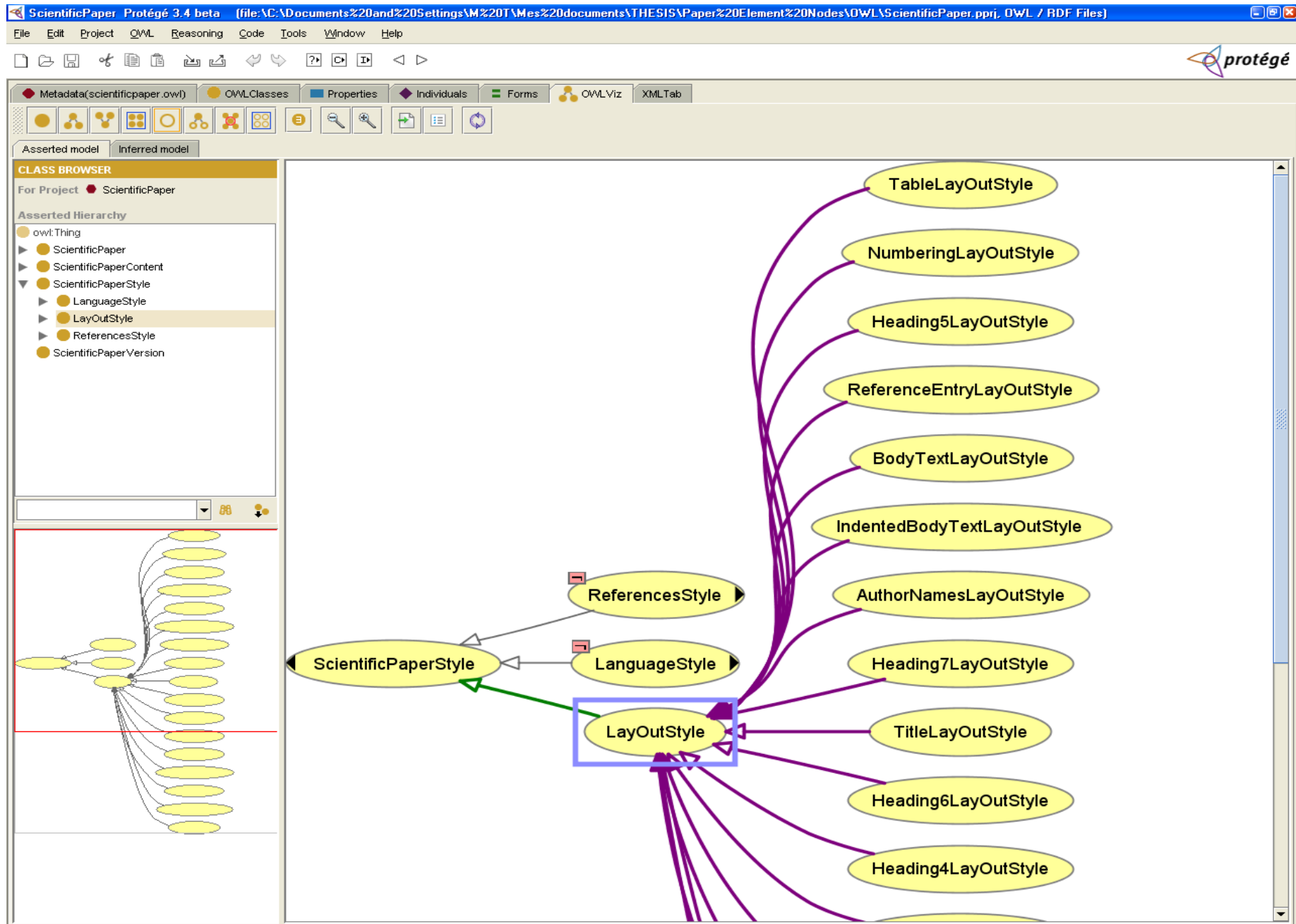


# Scientific Article Styles Class

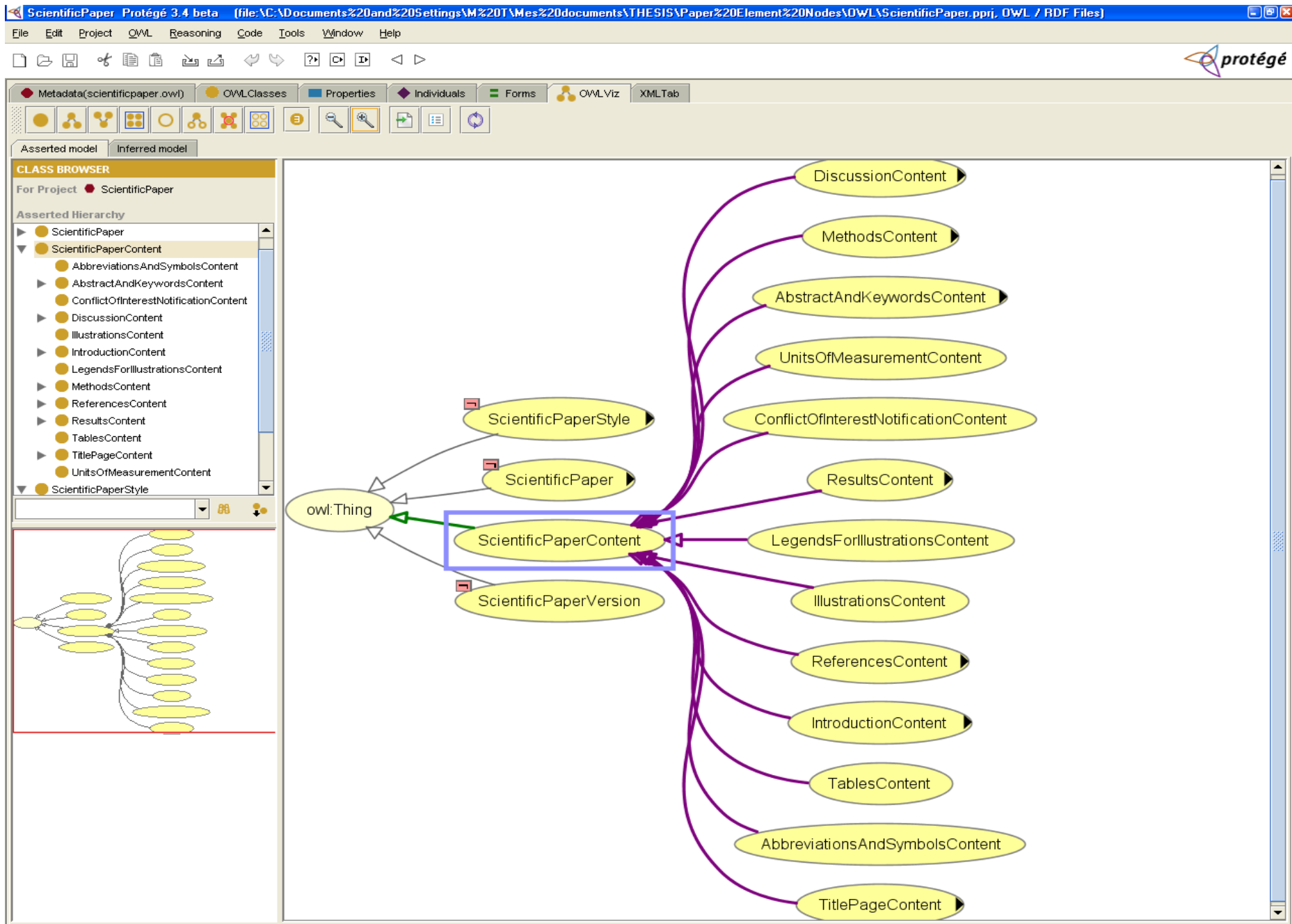




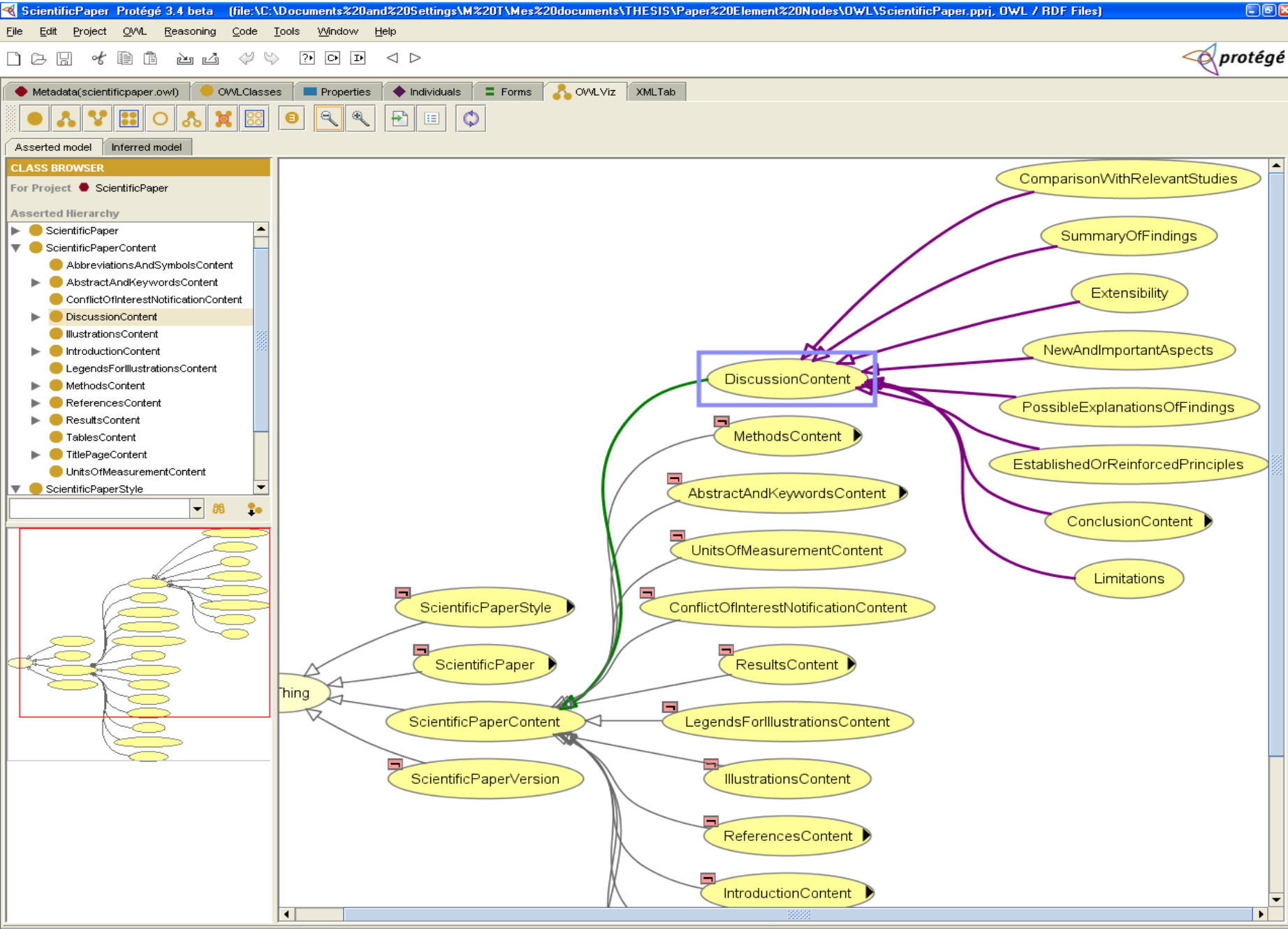
# Scientific Article Layout Styles



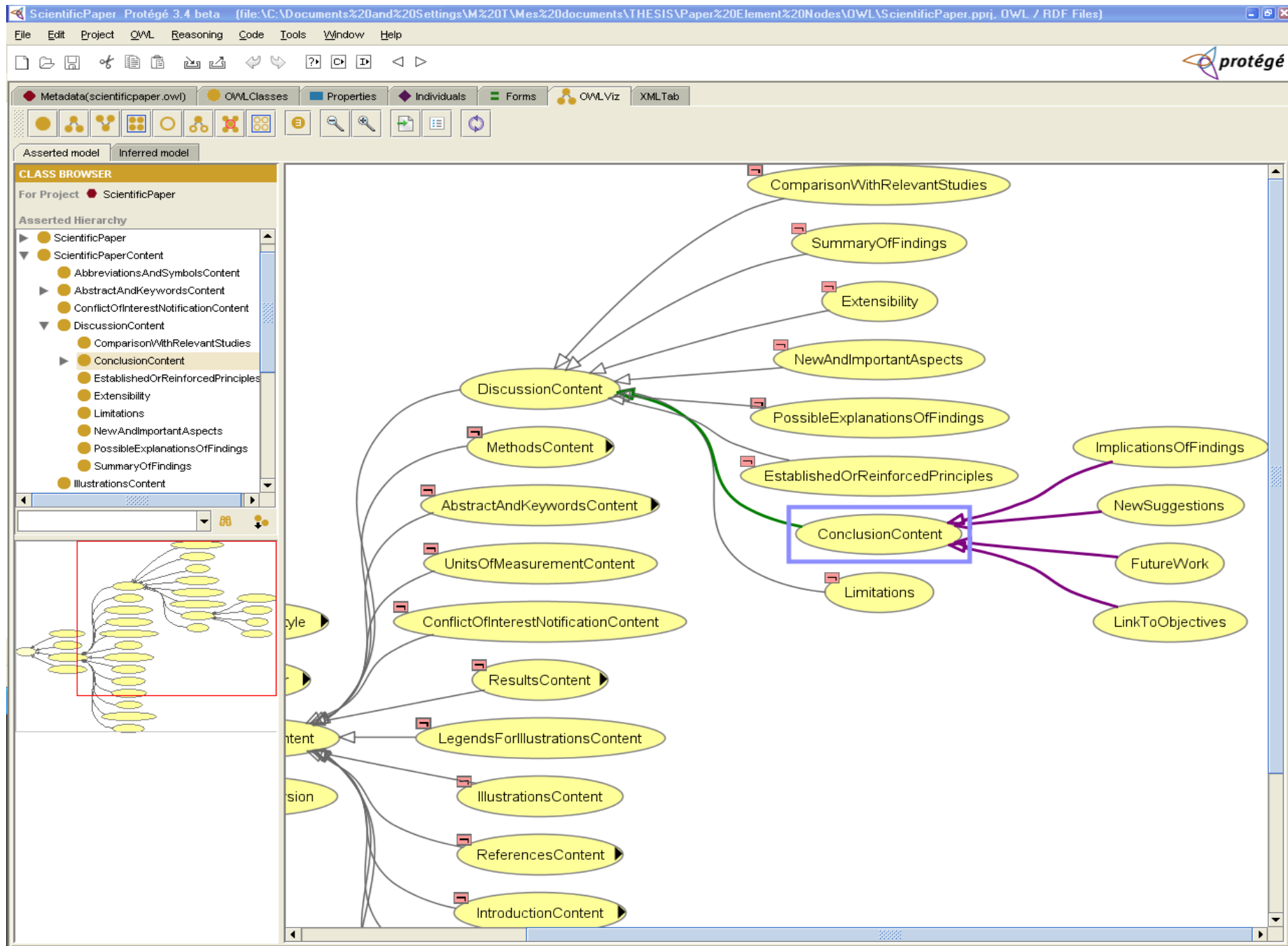
# Scientific Paper Content Class



# Discussion Content Class

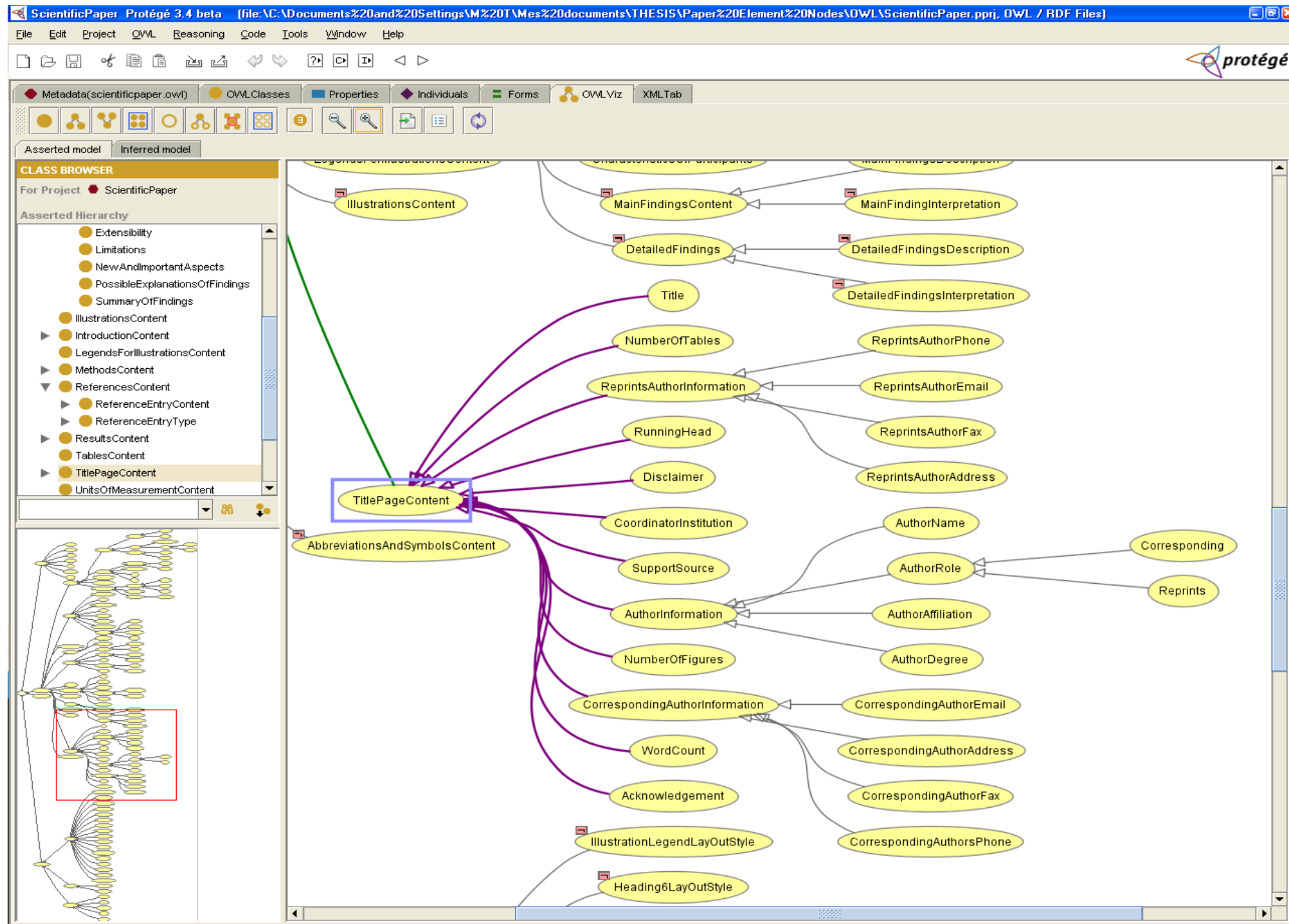


# Conclusion Content Class





# Title Page Content Class



# Exporting PEN in RDF

The screenshot displays the Protégé 3.4 beta interface with the 'ScientificPaper' ontology loaded. The main workspace shows a hierarchical ontology diagram with nodes like 'ScientificPaperContent', 'Abstract', 'TitlePageContent', and 'ResultsContent'. On the left, the 'CLASS BROWSER' shows the 'Asserted Hierarchy' for the project 'ScientificPaper', listing classes: 'owl:Thing', 'ScientificPaper', 'ScientificPaperContent', 'ScientificPaperStyle', and 'ScientificPaperVersion'. A dialog box titled 'OWL / RDF Files' is open, showing the export configuration:

- Project: documents\THESIS\Paper Element Nodes\OWL\ScientificPaper.pprj
- OWL file name or URL: ScientificPaper.rdf-xml.owl
- Language: RDF/XML

The dialog box includes 'OK' and 'Cancel' buttons. A small red box in the bottom-left corner of the main workspace highlights a portion of the ontology diagram.



# Details of PEN's OWL / RDF File

```
</owl:Class>
- <owl:Class rdf:about="#Limitations">
  - <owl:disjointWith>
    <owl:Class rdf:about="#SummaryOfFindings"/>
  </owl:disjointWith>
  - <owl:disjointWith>
    <owl:Class rdf:about="#NewAndImportantAspects"/>
  </owl:disjointWith>
  - <owl:disjointWith>
    <owl:Class rdf:about="#EstablishedOrReinforcedPrinciples"/>
  </owl:disjointWith>
  <owl:disjointWith rdf:resource="#ComparisonWithRelevantStudies"/>
  <owl:disjointWith rdf:resource="#PossibleExplanationsOfFindings"/>
  <owl:disjointWith rdf:resource="#Extensibility"/>
  + <owl:disjointWith></owl:disjointWith>
  <rdfs:subClassOf rdf:resource="#DiscussionContent"/>
</owl:Class>
+ <owl:Class rdf:about="#TableLayoutStyle"></owl:Class>
- <owl:Class rdf:about="#AuthorNamesLayoutStyle">
  - <owl:disjointWith>
    <owl:Class rdf:about="#Heading3LayoutStyle"/>
  </owl:disjointWith>
  - <owl:disjointWith>
    <owl:Class rdf:about="#IndentedBodyTextLayoutStyle"/>
  </owl:disjointWith>
  - <owl:disjointWith>
    <owl:Class rdf:about="#Heading1LayoutStyle"/>
  </owl:disjointWith>
  - <owl:disjointWith>
    <owl:Class rdf:about="#BodyTextLayoutStyle"/>
  </owl:disjointWith>
  <owl:disjointWith rdf:resource="#TitleLayoutStyle"/>
  <owl:disjointWith rdf:resource="#AuthorAffiliationsLayoutStyle"/>
  - <owl:disjointWith>
    <owl:Class rdf:about="#Heading7LayoutStyle"/>
  </owl:disjointWith>
  <owl:disjointWith rdf:resource="#Heading6LayoutStyle"/>
  <owl:disjointWith rdf:resource="#NumberingLayoutStyle"/>
  - <owl:disjointWith>
    <owl:Class rdf:about="#BulletLayoutStyle"/>
  </owl:disjointWith>
  - <rdfs:subClassOf>
    <owl:Class rdf:about="#LayoutStyle"/>
  </rdfs:subClassOf>
</owl:Class>
```

This format can be parsed by search engines but is not readable by office suits

# Making a PEN OpenOffice Template

The screenshot displays the OpenOffice Writer interface with a document titled "Massoud Toussi Paper Element Nodes V5.ott". The document contains several fields for a paper template, including title, author information, acknowledgments, keywords, word count, number of pictures and tables, abstract, background, objectives, and methods. A dialog box titled "Éditer les champs: Fonctions" is open, showing options for field types and formats.

**Document Fields:**

- <TITLE [STATEMENT: DESING]>
- <AUTHOR'S NAMES, DEGREES>
- <AUTHOR'S AFFILIATIONS>
- Corresponding author
- <CORRESPONDING AUTHOR'S NAME AND ADDRESS>
- Acknowledgments
- <ACKNOWLEDGMENTS>
- Keywords
- <KEYWORDS>
- Word count: <WORD COUNT>
- Number of pictures: <NUMBER OF FIGURES>
- Number of tables: <NUMBER OF TABLES>
- Abstract**
- Background**
- <BACKGROUND>
- Objectives**
- <MAIN AND SECONDARY OBJECTIVES>
- Methods**
- <METHODS>

**Dialog Box: Éditer les champs: Fonctions**

Type de champ	Format	Substituant
Substituant	Texte	Title [Statement: Desing]
	Tableau	
	Cadre	
	Images	
	Objet	

Annotation: Click to insert the title of your artic

Buttons: OK, Annuler, Aide, Macro...

# Content.xml file of the .ott file

```
- <office:document-content office:version="1.2">
  <office:scripts/>
  + <office:font-face-decls></office:font-face-decls>
  + <office:automatic-styles></office:automatic-styles>
  - <office:body>|
    - <office:text text:use-soft-page-breaks="true">
      + <text:sequence-decls></text:sequence-decls>
      - <text:p text:style-name="Title">
        <text:placeholder text:placeholder-type="text" text:description="Click to insert the title of your article here. It may have the form of
        "Statement : the design""><Title [Statement: Desing]></text:placeholder>
        </text:p>
      - <text:p text:style-name="Subtitle">
        <text:placeholder text:placeholder-type="text" text:description="Click to write authors' names and degrees"><Autho's names, degrees>
        </text:placeholder>
        </text:p>
      - <text:p text:style-name="Subtitle">
        <text:placeholder text:placeholder-type="text" text:description="Click to insert authors' affiliations"><Author's affiliations>
        </text:placeholder>
        </text:p>
        <text:h text:style-name="Heading_20_5" text:outline-level="5">Corresponding author</text:h>
      - <text:p text:style-name="Text_20_body">
        <text:placeholder text:placeholder-type="text" text:description="Click to enter the correspondig author's address"><Corresponding
        author's name and address></text:placeholder>
        </text:p>
        <text:h text:style-name="Heading_20_5" text:outline-level="5">Acknowledgments</text:h>
      - <text:p text:style-name="Text_20_body">
        <text:placeholder text:placeholder-type="text" text:description="Write the name of anyone who worked on the study but not fulfilled
        the criteria of authors."><Acknowledgments></text:placeholder>
        </text:p>
        <text:h text:style-name="Heading_20_5" text:outline-level="5">Keywords</text:h>
      - <text:p text:style-name="Text_20_body">
        <text:placeholder text:placeholder-type="text" text:description="Enter up to 10 keywords. It is preferred to use Medical Subject
        Headings (MeSH) keywords."><Keywords></text:placeholder>
        </text:p>
      - <text:h text:style-name="Heading_20_5" text:outline-level="5">
        Word count:
        <text:placeholder text:placeholder-type="text" text:description="Click to enter the word count (from introduction to
        conclusion)"><Word count></text:placeholder>
        </text:h>
      - <text:h text:style-name="Heading_20_5" text:outline-level="5">
```



# Discussion

- ▶ We created an OpenOffice template which keeps both the layout and the semantic structure of a scientific paper
- ▶ This method as well as the PEN ontology can be generalized to domains other than biomedical sciences
- ▶ Critics:
  - ▶ The process of transforming OWL to a template is not automatic, but it is possible to do so.
  - ▶ The order is not specified in OWL ontologies. (adding *follows* / *isFollowedBy* property or add *hasOrder* / *isOrderOf* ?)



# Conclusions

- ▶ Similar ontologies can be created for other types of documents, such as hospitalization reports, film scripts, essays,...(and why not love letters?).
- ▶ Perspectives:
  - ▶ A new generation of office suits can be imagined which will help organising ideas more efficiently
  - ▶ Search engines can be enabled to parse this file for advanced semantic research in scientific papers, so that:
    - ▶ Searching articles about an OWL ontology of birds (including owls) would no longer be messy!





Thank you

[massoudtoussi@gmail.com](mailto:massoudtoussi@gmail.com)  
CloseClinical.org