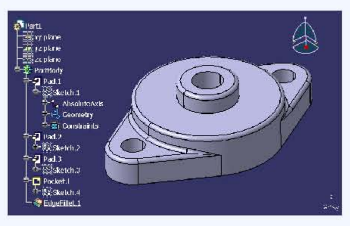
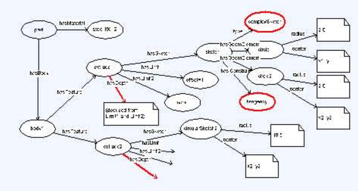


Relational Data Mining Applied to Virtual Engineering of Product Designs



Monika Žáková[1], Filip Železný[1], Javier A. Garcia-Sedano[2], Cyril Masia Tissot[2] and Nada Lavrač[3,4]

[1] Czech Technical University, Prague, Czech Republic, [2] Semantic Systems, Av. Del Txoriherri 9, Derio, Spain

[3] Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia, [4] University of Nova Gorica, Nova Gorica, Slovenia

Semantic Virtual Engineering

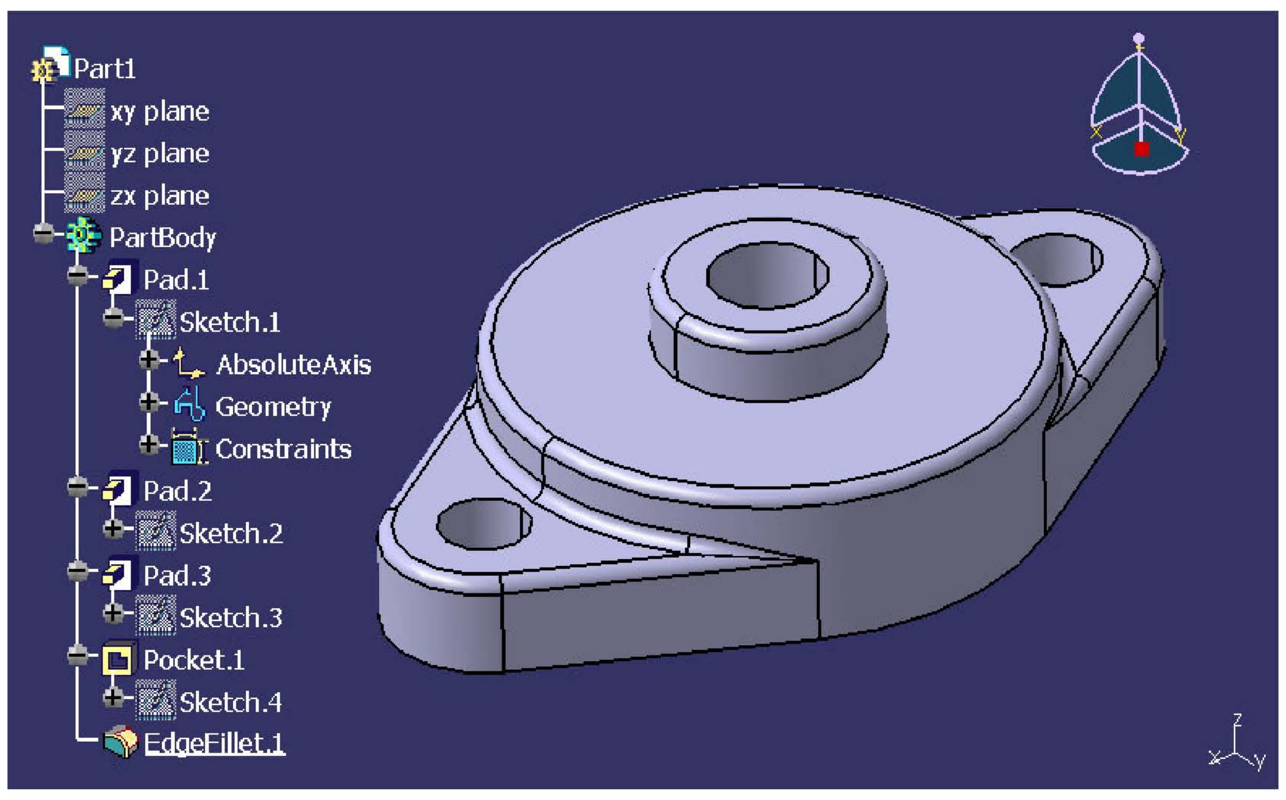
Project Objectives

The SEVENPRO project aims at developing technologies for

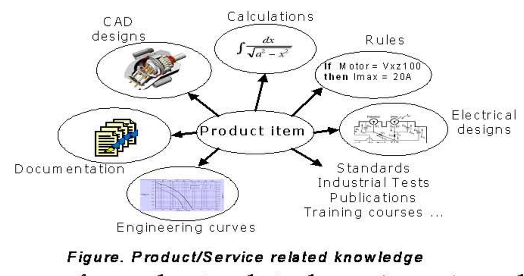
- **mining of product engineering knowledge from multimedia repositories** and
- semantically enhanced 3D interaction with that knowledge in integrated engineering environments through **virtual reality.**

CAD designs, documental repositories and enterprize-resource-planning databases will be the main data&knowledge sources supported.

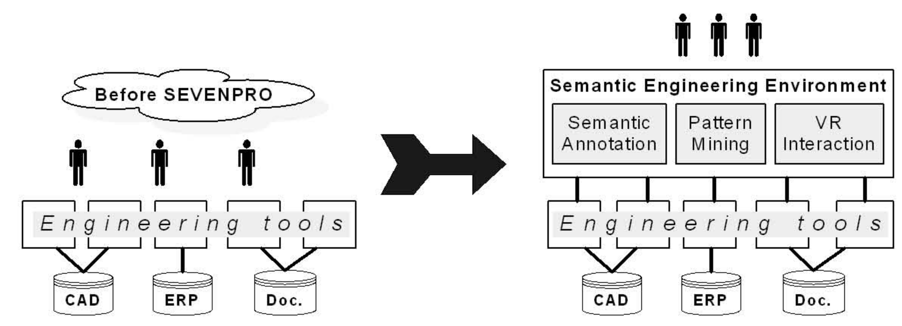
The project brings together the so far distinct worlds of 3D design data and other representation formats of engineering knowledge .



Example of a CAD design of two bolt flange includig commands history. A design is obtained by successive applications of CAD commands, such as *extrusion*, *rotation*, etc., which are mutually parametereically related. Various command sequences may lead to the same design, but differ greatly in quality respects, such as complexity, reusability, etc.



Various types of product-related engineering data to be integrated by the project.



The integration will allow the engineer, for example, to query about "parts of type clamp [*itemFamily*], with more than 6 holes [*Geometry*], set in order later than November/2004 [*Management*], compliant with ISO-23013 [*Documentation*]". This is not possible with current information systems.

Designs Annotation

The information available in CAD files and other data sources is formalized and integrated by means of semantic annotation based on ontologies.

Semantic annotoation of CAD designs

- is generated automatically from the commands history available via the API of CAD tools
- is based on a CAD ontology developed in SEVENPRO and available in RDF format

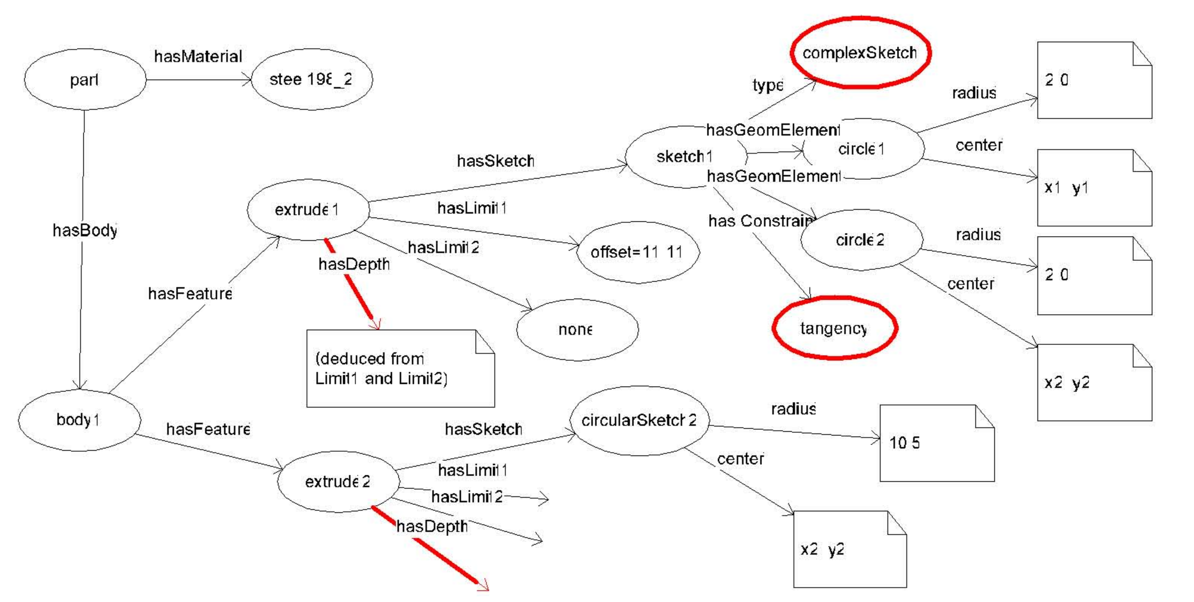
Annotation including ontology of CAD items and axioms defining core relations is automatically translated into Prolog.

Challenges to ILP

There are three main challenges for ILP due to ontolgies in the background knowledge:

- hierarchies of concepts induced by *subclassOf* relation
- hierarchies of relations induced by *subpropertyOf* relation
- representation conversion between Prolog and other knowledge representation languages

Therefore an efficient handling of term taxonomies has to be integrated into the RDM system. To exploit the subproperty relation directly, the system would deal with taxonomies of predicates. Also, to make the induced rules available to the semantic server, conversion to SWRL is considered.



Part of a semantic annotation of a two_bolt_flange. It corresponds to the commands history used to design two bolt flange using a CAD tool. The edges in the graph represent the CAD features or specific properties of those features. Appart from objects and properties occurring in the original CAD design, some derived objects and properties have been added to the annotation such as *hasDepth*, *complexSketch* and *tangency*.

Relational Mining from Product Designs Data

Mining Engineering Designs

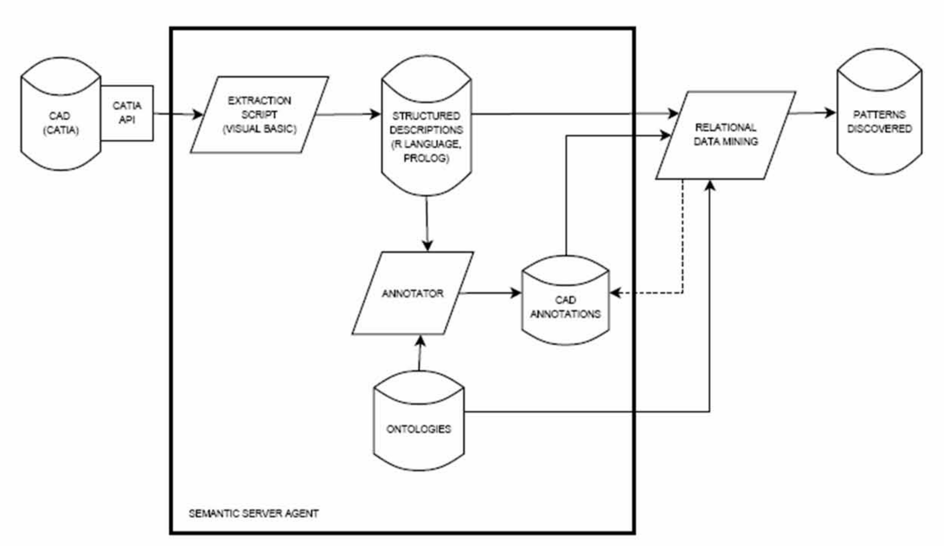
Discovery of interesting design patterns from CAD repositories is an expected aid to reusing design knowledge. Engineering designs contain **implicit expert knowledge** the data mining algorithms will help to **reveal and make explicit**. Such obtained knowledge can be reused as corporate design standards or recommendations: to support the work of engineers (reusability), to check pattern compliance of new designs (quality checking), to teach novel engineers on how to design specific parts (training).

Our Approach

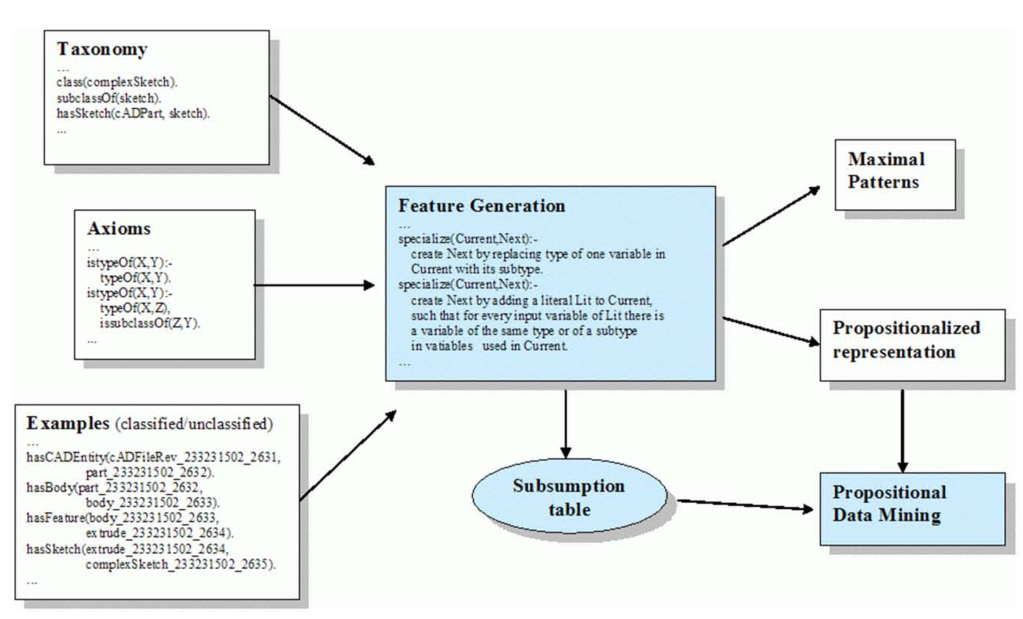
Our approach is based on the sorted refinement operator, i.e. a subsumption relation combining classical θ-subsumption with taxonomies on terms and predicates.

Propositionalization enables us to exploit the whole variety of available data mining algorithms. Propositionalized representation of classified relational data is generated by constructing first-order features. During the feature generation a table of mutual feature subsumptions is maintained. This subsumption is exploited in propositional pattern search, which prunes any conjunctions of a subsumer with its subsumee and specializes a conjunction not only by extending it, but also by replacing an included feature with its subsumee.

Finding maximal patterns For non-classified data, find maximal patterns of some maximal length covering the minimum set amount of examples.



The input data and output of the RDM component. Processes generating the main input data to the RDM component are: an extraction script producing a structured representation of the original CAD data in a minable representation format, and annotator linking items of the CAD data onto the predefined CAD ontology.



Schema of the RDM component. A training example is an annotation of a CAD design. It consists of binary ground facts and corresponds to a rooted tree, where the root entity is the CAD file containing the design. The subsumption table created during generation of first-order features is then used to boost the performance of the propositional data mining algorithm.

Results and Ongoing Work

In a proof-of-concept experiment, the system was tested on a database of 35 CAD designs. One design usually contains around 100 properties. It is not feasible to generate features with tens of literals, therefore the maximum length of a feature was set, based on maximum depth and highest occurrence of one property in the dataset. An exemplary maximal pattern is below:

f(X1:cADFileRevision) = hasCADEntity(X1:cADFileRevision,X2:cADPart),hasBody(X2:cADPart,X3:body),hasFeature(X3:body,X4:extrude),…,hasFeature(X3:body,X7:extrude),hasSketch(X7:extrude,X8:circularSketch),hasGeomElement(X8:circularSketch,X9:circle).

Variable

Term class subclass of "sketch"

Currently we are struggling to improve efficiency by use of graph search techniques for feature generation. For principled exploitation of hierarchical background knowledge, we are also considering approaches using hybrid languages such as AL-log or using a more complex description language such as ψ-terms or antecedent description grammars.

This research is supported by **SEVENPRO: Semantic Virtual Engineering for Product Design**, project IST-027473(2006-2008) funded under 6th Framework Programme of the European Commission.