# Incorporating Cluster-Based Relationships in Web Rule Language

**Mala Mehrotra**
**Pragati Synergetic Research Inc.**
**Cupertino. CA**
**mm@pragati-inc.com**
**http://www.pragati-inc.com**

## 1 Introduction

The Semantic Web vision requires rule-based systems to express the knowledge necessary for extracting and utilizing distributed information on the web. As the semantic web grows, it is important that we facilitate the maintenance, reuse and interoperability of axioms by allowing higher-level knowledge about them to be expressed via more expressive rules.

In the course of analyzing a wide variety of knowledge-based systems, Pragati has found many types of cluster-based relationships that can enable analysts and developers to comprehend, maintain, and reuse such systems more effectively. Pragati's cluster-based analysis of numerous knowledge-based systems through its Multi-ViewPoint Clustering Analysis (MVP-CA) Tool [Mehrotra & Bobrovnikoff 2002] has demonstrated that clustering can expose a wide variety of relationships between formally represented concepts. These include templatization and refactoring opportunities, exposition of idioms, such as inverse rules, idempotent rules, and potential inter-system mappings.

Pragati's existing tool suite provides statistics- and heuristics-based clustering on axioms in knowledge bases that enables sophisticated analytical capabilities on the properties of the clusters of axioms. Higher-level relationships based on the clusters have been the domain of human analysts, and have been generally expressed as written reports and hand-generated diagrams.

*It is our position that any rule language for the semantic web should have the expressivity and the extensibility to represent diverse metadata to encapsulate such findings.* This position paper presents some representative results from Pragati's high-level clustering analysis that needs to be expressed in the W3C rules being designed for interoperability, reuse and maintenance.

## 2 Representing Cluster-Based Analysis Results

Pragati's technology consists of an integrated suite of cluster-based cognitive assistance tools based around Pragati's core capability of automatically grouping concepts through the MVP-CA tool. It facilitates analysis of rule based systems by clustering the rules into clusters that share significant common properties. It exposes rule-based system developers to the usage of the rules by clustering the information units that exercise/use the system in a given practical context. In its current state, MVP-CA tool can accept many different types of knowledge bases or decision support systems and expose conceptual islands or clusters of concepts in the vicinity of a seed concept. These vicinity concepts aid the user's comprehension of knowledge bases and lead to "fortuitous" reuse opportunities—even when the source knowledge bases are unfamiliar to the user [Hayes et.al 2005**].** In this section we will present some representative examples of issues

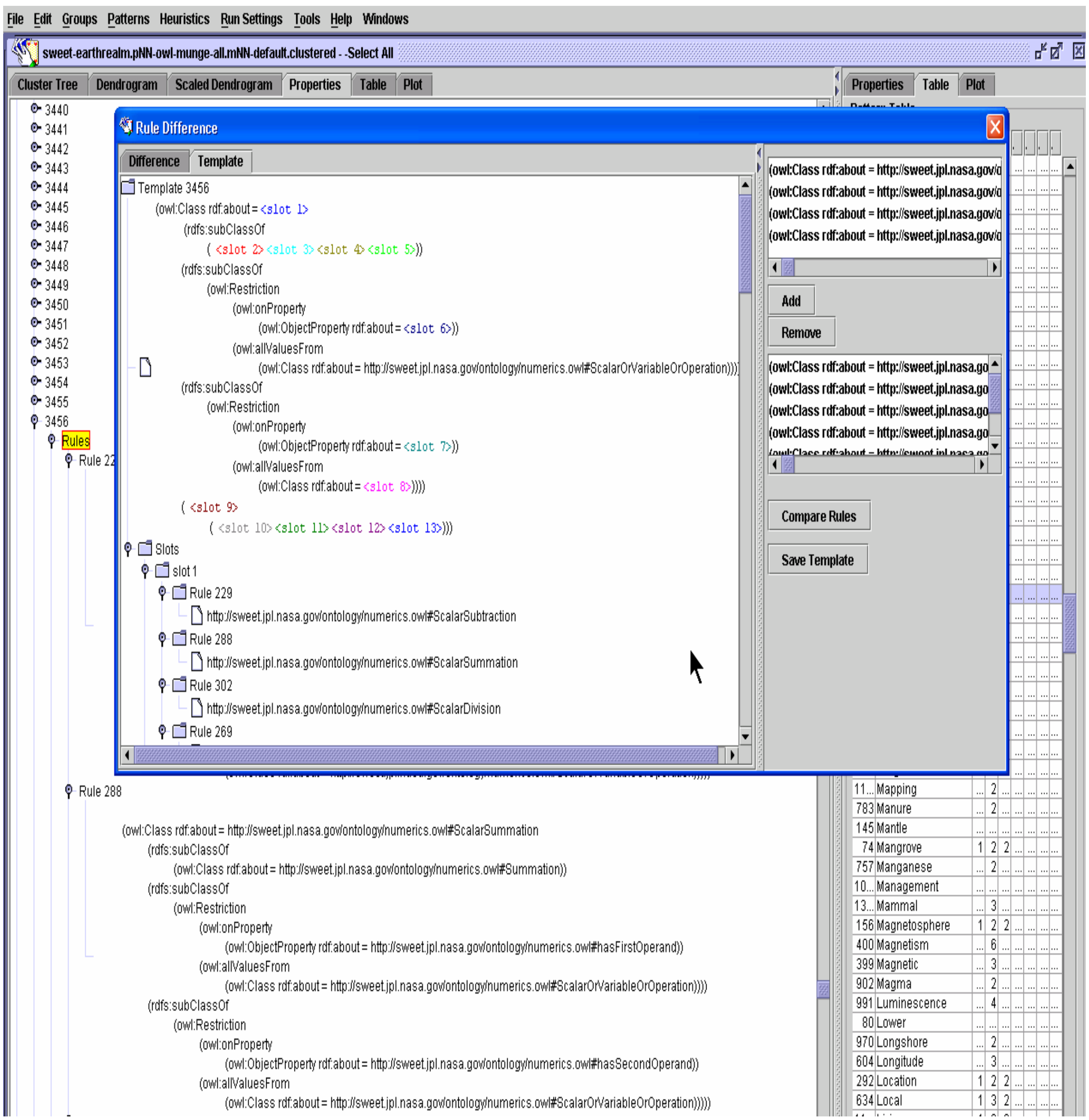


*Figure 1: Snapshot of the MVP-CA Tool – Template formed from JPL's* SWEET Earthrealm OWL Ontology

found through the MVP-CA tool which can help influence the design of the web language of future.

## 2.1 Templatization of Usage Patterns

Clusters of structurally similar axioms signal prototypical usage patterns for concept(s). Identification of such regions in the software can be a valuable guide for future extensions of the knowledge base, as well as provide us opportunities for identifying generic, reusable regions in the software [Mehrotra et.al 1999]. Exposing common functionalities, as well as overlapping contexts across clusters of axioms, reveals opportunities for creation of higher order predicates for axioms, such as macro predicates in Cyc, and templates for groups of axioms [Mehrotra 2002]. Often the need for these higher-level abstractions can be revealed only ex post facto, that is, after there are enough assertions in the knowledge base to warrant the formation of a higher order axiom. These important knowledge representational issues can impact the long-term utility and quality assurance of the knowledge base.

In Figure 1, we present a snapshot of the templating infrastructure from the MVP-CA tool showing a representative cluster from JPL's SWEET Earthrealm knowledge base written in OWL DL. The template is formed automatically in the overlaid window by performing rule differencing operation on a selected rule set. The cluster selected in Figure 1 shows how different types of scalar operations, such as, *ScalarSummation, ScalarSubtraction, ScalarMultiplication and ScalarDivision* are defined OWL-DL by placing similar value restrictions on the class *ScalarOrVariableOrOperation*, and similar property restrictions on *hasFirstOperand* and *hasSecondOperand.* These templates can be flagged and documented in the current MVP-CA infrastructure. In our experience with various types of knowledge bases, opportunities for template formation arise repeatedly, regardless of the domain or the representation language. Capturing these templates in a succinct rule-representation language will allow them to become easily available for search, retrieval and reuse across knowledge bases in the larger context of the semantic web.

## 2.2 Refactoring

In many cases, we have found portions of knowledge-based systems that could benefit from various types of refactoring. Figure 2 shows a dependency diagram view of two clusters and the mostly-overlapping term sets that they use. The domain rules in this case are from the IMMACCS (Integrated Marine Multi-Agent Command and Control System) Clips system and they represent the various situations when the object between the weapon and target trajectory could be a rotary wing or a blocking building. In this case, the rules in the clusters are parallel, and it would be useful to refactor both the rules and the ontological definitions they use [Mehrotra & Bobrovnikoff 2001].

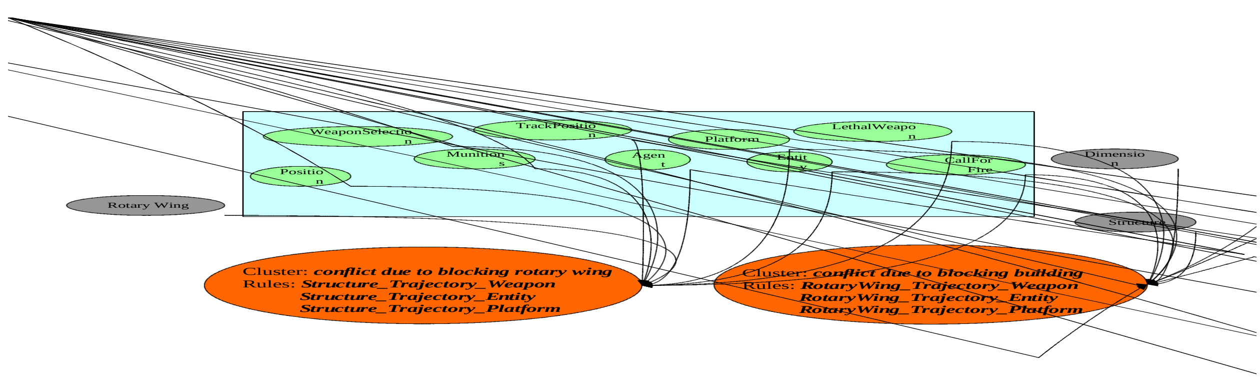


*Figure 2:* Refactoring the concept of "conflicts" from IMMACCS Fires agent

However, while it is possible to model refactoring at several different levels of detail, a good starting point would be to describe the starting set of entities and the ending (i.e., transformed)

```
(#$implies (#$and            (#$subSituationTypes ?COA ?TASK)            (#$increasesSitType-G
    (#$evaluatesDimensionOfCOA ?COA                                           #$mitigateRiskFactorsI
```

set of entities. At this level, it would be necessary to represent the entities first as common subregions in a rule-language and have transformations on those rules to achieve the refactoring effect.

### 2.3 InverseConcepts

In the third example, shown in Figure 3, we see that there are two legitimate ways of expressing the *mitigateRiskFactorsInCOA* and *Effective/IneffectiveForPurpose.* On close analysis these Cyc [Cycorp] axioms can be recognized as near inverses of each other [Pool et.al 2003]. The user can create an association between the rule entities and label the association "near inverse of". Future queries that turn up either one of the rules should also turn up the companion opposite rule so that the authors are aware of the various legitimate representations. This requires us to define an *inverse rule* construct in the web rule language just as we have for inverse concepts.

## 3 Bibliography


CYCORP http://www.cyc.com/

P. Hayes, T. Eskridge, M. Mehrotra, D. Bobrovnikoff, T. Reichherzer, R. Saavedra. (2005) *COE: Tools for Ontology Development and Reuse.* Submitted to 2005 International Conference on Intelligence Analysis, May 3-5, 2005, McLean, VA.

M. Mehrotra. (2004) Semantic Mediation Tools for Interoperability. 6th Annual ONR Workshop on Collaborative Decision Support Systems. Sept 8 - 9, 2004. Quantico, VA. pp: 261-269.

M. Pool, K. Murray, J. Fitzgerald, M. Mehrotra, R. Schrag, J. Blythe, J. Kim, H. Chalupsky, P. Miraglia, T. Russ, & D. Schneider. (2003} *Evaluating SME-Authored COA Critiquing Knowledge*. K-CAP 2003, October 23-25, Sanibel Island, Florida.

Mehrotra M.. (2002) Ontology analysis for the Semantic Web. Ontologies and the Semantic Web: Papers from the AAAI Workshop (Technical Report WS-02-11). Edmonton, Canada: AAAI Press.

Mehrotra, M., & Bobrovnikoff, D. (2002). Multi-viewpoint clustering analysis tool. Proceedings of the Eighteenth AAAI Conference (pp. 1006-1007).Edmonton, Canada: AAAI Press.

Mehrotra M., & Bobrovnikoff, D. (2001). Multi-Viewpoint Clustering Analysis (MVP-CA) Technology for Analysis of the Integrated Marine Multi-agent Command and Control System (IMMACCS). Proceedings of ONR Decision-Support Workshop on Continuing the Revolution in Military Affairs. June 2001, Quantico, VA, CDM Technologies, San Luis Obispo, CA.

Mehrotra et al. (1999) M. Mehrotra, S. Alvarado and R. Wainwright. Laying a Foundation for Software Engineering of Knowledge Bases in Spacecraft Ground Systems. Proceedings of FLAIRS99 Conference. Special Track on AI Applied to Spacecraft Autonomy, Florida.

Mehrotra M. and Wild C. (1995) Analyzing Knowledge-Based Systems using Multi-Viewpoint Clustering Analysis. Journal of Systems and Software, 29:235-249, 1995.