Expozé Tool Suite

Mala Mehrotra
President
Pragati Synergetic Research Inc.

Contact-Info: mm@pragati-inc.com
408-910-4115 (Cell)
650-625-0274(Office)
408-861-0939

Keywords: knowledge discovery, interoperability, reuse, clustering, mapping.

Pragati's Expozé tool suite, a prototype being developed with ONR SBIR Phase II funds for the project entitled "A Clustering-Based Infrastructure for the Reuse of Ontologies" (Solicitation Topic Number: N03-106), is an integrated suite of cognitive assistance tools, based around the core capability of clustering concepts which have structural and semantic similarities in their descriptions. Pragati Inc. has focused on this technology for the last 13 years developing proprietary heuristic algorithms for clustering as well as devising cognitive aids for exposing the meaningful results from clustering. Expozé is a human-machine interface tool that eliminates/greatly reduces the cognitive overload found in currently used analysis tools and manual methods [1]. It suitably abstracts, structures and clusters knowledge in various information systems such as, expert systems, knowledge-bases, databases, and stylized natural-language text, in a manner that facilitates the following goals:

- **Comprehension**
  High-level understanding of the software system, both from hierarchical (detail to abstract) and orthogonal (contextually distinct) perspectives.
- **Knowledge Reuse**
  Recognition of reuse opportunities in the systems through semi-automated component and template formation.
- **Mapping**
  Automated mapping and alignment support for multiply authored systems by drawing attention to common/overlapping contexts in software.
- **Quality Assurance**
  Long-term quality assurance of software systems by exposing "infelicitous" knowledge entry patterns, inconsistent and redundant concepts for meaningful knowledge repair on the information system.
- **Information Extraction**
  Ontological engineering support for building ontologies and taxonomies by exposing salient concepts from a legacy information system and suggesting appropriate placement of domain terms in an ontological hierarchy.

  Based on the above objectives we have the following interfaces built into our tool suite:

  1. **Query and Analysis Interface is composed of the following five views:**
     - *Vicinity Concepts View*
       High-level view of the surrounding context of a term in its cluster
     - *Cluster View*
       More detail on contextual and hierarchical cluster relationships
     - *Term Relationships View*
       Co-occurrence relationship with other terms in the cluster
     - *Entities View*
       Actual description and text of entities that form the basis for clustering
     - *Graphical View*

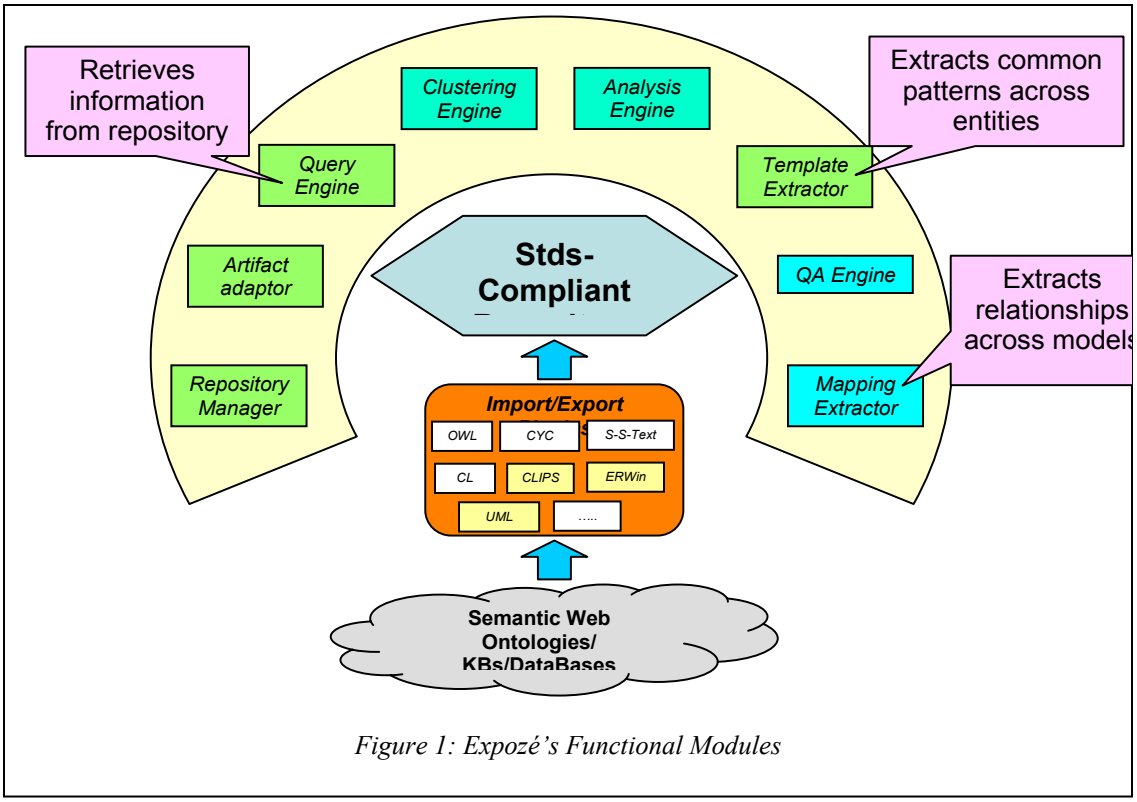

Figure 1: Expozé's Functional Modules

- Graphical depiction of inter-concept relationships as defined in the information source

2. **Reuse Interface**
   - *Template creation*
     Capture similarity of recurring knowledge entry patterns as template with open slots
   - *Template search*
     Retrieve previously stored templates from the Expozé repository
   - *Template reuse*
     Adapt templates in the repository for forming new axioms/entities by providing new values for the open slots
3. **Mapping Interface**
   Identification of potential mappings based on:
   - *Syntax of information sources*
   - *Intended semantics in the source ontologies based on context of usage of concepts*
   - *Rich relationships across ontological concepts in the source ontologies*

Expozé has been used to analyze a large number of knowledge-based sources in different representation-languages and from a variety of domains[2,3,4,5,6]. Figure 1 presents a high-level functional view of Expozé's various components. Different types of source datasets, in the form of axiomatized knowledge-bases (CL, OWL, CLIPS, CycL, etc.), databases (ErWin relational database, SQL-based database) or stylized natural language can be ingested into Expozé's language-neutral repository. The clustering engine clusters concept terms from ontologies that have similar formal descriptions. Datasets can be clustered one at a time or in a combined fashion. The concepts that cluster together often do not have direct formal relationships defined across them in the ontology; these are implicit connections found by noting structural similarities between the axioms/entities which contain them and, recursively, the similarity of terms surrounding the definitions of the concepts. Exposing such *vicinity concepts* to the user provides a quick, intuitive "peek" into likely similarity of intended meanings of concepts in different ontologies and aids exploratory ontology building, and can lead to "fortuitous" re-use opportunities. The analysis engine aids the discovery of such regions. Figure 2 provides a snapshot of the query engine user interface. Mapping extractor currently allows ingestion of multiple data sources in a tagged form and exposes clusters that

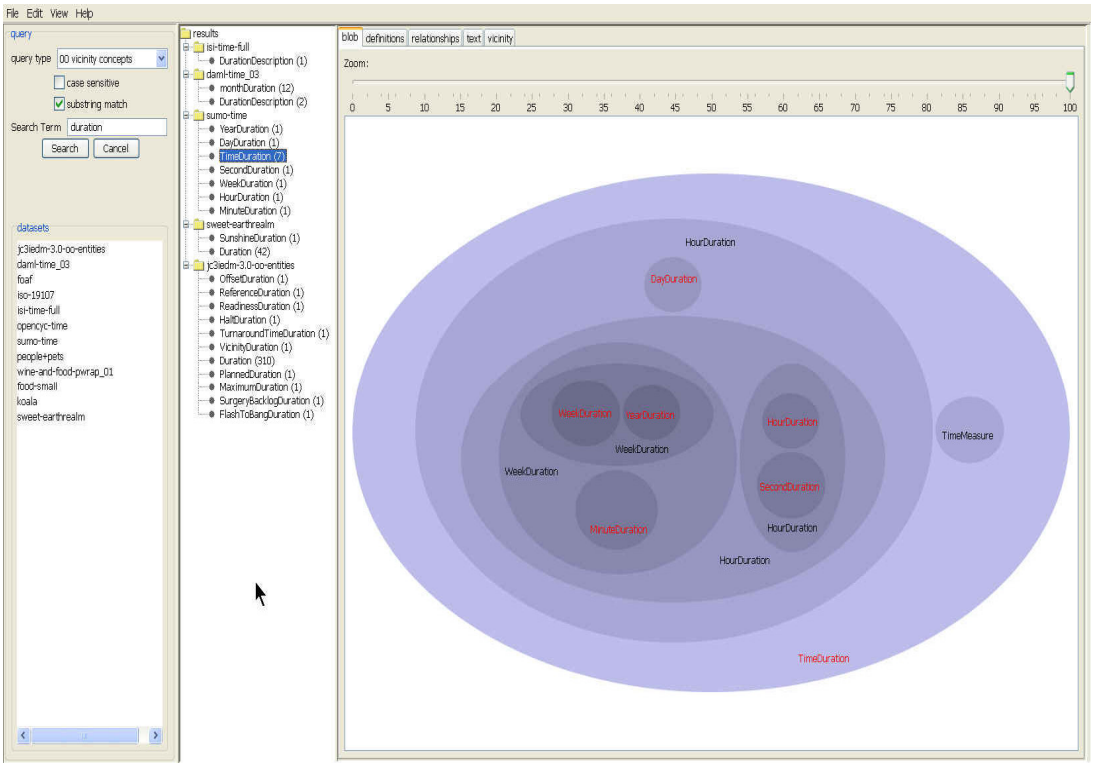

Figure 2: Expozé's Query Interface

contain the similar concepts across them. Disambiguation of suitable candidates for mapping is carried out in a human machine collaboration mode. We are currently building web-services API that will provide access to all the module functionalities. We are also positioned to plug into other frameworks such as authoring and mapping. We have demonstrated this capability in our collaboration with COE (Collaborative Ontology Authoring Environment) of IHMC (Institute for Human & Machine Cognition) [1] and are looking for similar collaborations with Protege's Prompt. The representation-neutral aspect of Expozé's repository positions the tool suite to be utilized across multiple standards-compliant domains. We are participating in standards efforts like XMDR[7] and RIF[8] to influence their design to take advantage of cluster-based information.

References:

1. P. Hayes, T. Eskridge, M. Mehrotra, D. Bobrovnikoff, T. Reichherzer, R. Saavedra. *Collaborative Knowledge Capture in Ontologies.* Knowledge Capture Conference Oct 2-6, 2005, Bannf, Canada.
2. M. Mehrotra. *Ontology Analysis for the Semantic Web*. AAAI-02 Workshop on Ontologies and the Semantic Web. Edmonton, Canada. July 28-29, 2002.
3. M. Mehrotra. *Semantic Mediation Tools for Interoperability.* 6th Annual ONR Workshop on Collaborative Decision Support Systems. Sept 8 - 9, 2004. Quantico, VA. pp: 261-269.
4. M. Mehrotra and D. Bobrovnikoff. *An Analysis Tool Suite for Management and Reuse of Ontologies.* ONR Workshop Series on Collaborative Decision Support Systems. Sept 18 -19, 2002. Quantico, VA. pp: 241-254.
5. M. Mehrotra and D. Bobrovnikoff. *Multi-ViewPoint Clustering Analysis (MVP-CA) Technology for Analysis of Integrated Marine Multi-Agent Command and Control System (IMMACCS).* Proceedings of ONR Decision-Support Workshop on Continuing the Revolution in Military Affairs (RMA). June 5-7, 2001. Quantico Marine Base, Ft. Quantico, VA.
6. RKF PI presentations http://cerberus.cyc.com/RKF/
7. http://hpcrd.lbl.gov/SDM/XMDR/presentations/
8. http://www.w3.org/2005/rules/wg/wiki/Supporting_the_Reuse_of_Rules