A Pilot Project for Knowledge Extraction

Our pilot project for Knowledge Extraction is based on previous work in which we assembled the historical literature (1800s-2016) for all named organisms. The reasoner interprets these assertions at appropriate levels of multiple levels of imprecise terms at verifiable confidence intervals. We have subsequently developed this vocabulary into a thesaurus of phenotypic terms, which will ultimately provide a novel way to navigate and bridge multiple classification systems.

A Hybrid Knowledge Framework

We developed software to extract pertinent feature domain vocabulary from this corpus, resulting in the discovery of over 20,000 candidate terms used in descriptions of the 12,123 species of Bacterium and Archaea (Int. Bact., December 1, 2016). We have subsequently developed this vocabulary into a thesaurus of phenotypic terms, which will ultimately conform to the W3C SKOS-XL semantics and provide a link between the language of microbiology, the semantic web and our existing tools and services.

The thesaurus was developed in parallel with a formal ontological schema, which supports inference from observations of an organism under a set of environmental constraints. The ontology provides a set of generic axioms to implement rule and constraint templates using these complex terms.

While developing these resources, we discovered a novel method for establishing semantic equivalence between concepts. This enables precise, consistent, verifiable reasoning over imprecise terms at multiple levels of abstraction [1].

Our thesaurus model complements ontology development by deconstructing ambiguous language into precise concepts for reasoners. Our approach employs semantic and semiotic methods to:

• represent information at multiple levels in concept hierarchies
• “slice” and aggregate concepts to represent information consistently for ambiguous language and reasoners
• provide multiple entry points for information (term, concept, data)
• provide attachment points for reasoning over rules and axioms

We are developing a standards-compliant semantic knowledge model to support abstract reasoning and predictive modeling.

Capabilities & Commercial Offerings

We have already developed a number of novel software components that overcome specific technical barriers in terminology management, text mining, Information Extraction, knowledge transformation, name-entity recognition, document classification and annotation. Each component is implemented using W3C standards and recommendations (RDF/S, OWL2, SKOS-XL, XML, XSL, SPARQL, OWL RL, DO COKS) and commercially-compatible FOS frameworks (Java, Apache, PostgreSQL, Virtuoso, Jena/ARQ, SPARQL Reasoner). We are integrating these components into a single software suite that supports a variety of Machine Learning and reasoning needs.

In the text-mining field, we have demonstrated our ability to reverse-engineer the diagnostic phrases that human indexers use to classify large corpora of technical documents, and to measure both the quality of previously-annotated documents and the cohesion of individual document classes. Our software provides a novel way to navigate and bridge multiple classification systems.

Military Applications

Our company has developed a novel approach to knowledge modeling that bridges the gap between human and machine understanding of abstract concepts. This technology enables abstract reasoning and sharing of Mixed-Precision Information that can benefit Human/Autonomous System Interaction and Collaboration (HASIC) through bi-directional human-agent interaction.

Benefits of Our Approach

• Common understanding by both human and machine.
• Compliant with existing ontology and thesaurus standards.
• Preserves precision of the original data.
• Utilizes data at appropriate levels of abstraction.
• Accommodates conflicting and ambiguous information.
• Supports defensible inference of relations among concepts.
• Provides exact explanations for reasoning.
• Integrates data from multiple sources.

Company Information

NamesforLife, LLC is a privately held company based in East Lansing, Michigan. It was founded in 2004 to create commercial applications at a developing technology at Michigan State University. The company provides taxonomic and analytical services, software and technology licensing for the publishing industry, life sciences research, commercial and federal laboratories.

For more information, please visit NamesforLife’s website or contact us at 517-410-0525.

4227 (Polymerization Catalyst)

Enzyme Knowledge for Human and Machine Understanding

Our semantic equivalence method integrates observational data from multiple sources to infer abstract knowledge and ambiguity and detects conflicting observations prior to resolving to labeled ontology concept identifiers suitable for reasoning.

Discover and Label New Concepts

Our self-organizing business entities entities based on semantic (term) or semiotic (concept) similarity, discovering previously unknown entity groups and entities.

Multiple Classifications with Explanations

Reverse-engineer human analyses, discover errors and perform multiple-classifications automatically.

Visible Group Quality

Multi-dimensional analysis of shared concepts within a group provides a means to visualize group cohesion.

References:


