Semantic Web Infrastructure for Fungal Enzyme Biotechnologists

Christopher J. O. Baker[1], Arash Shaban-Nejad[1], Xiao Su[1], Volker Haarslev[1] Greg Butler[1]

[1] Department of Computer Science and Software Engineering, Concordia University,
1455de Maisonneuve Ouest, Montreal, Quebec, Canada, H3G 1M8
{baker, arash_sh, x_su, haarslev, gregb}@cs.concordia.ca
http://www.cs.concordia.ca/FungalWeb/

Abstract

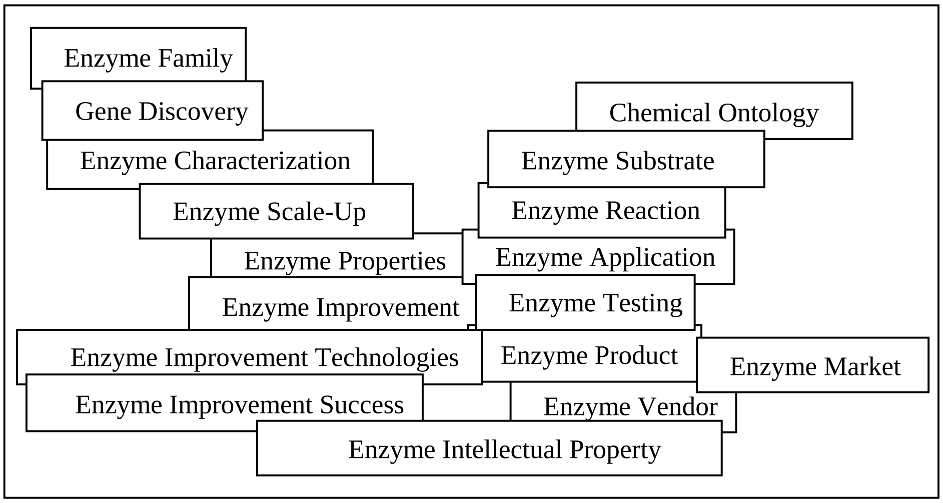
The FungalWeb Ontology aims to support the data integration needs of enzyme biotechnology from inception to product roll. Serving as a knowledgebase for decision support, the conceptualization seeks to link fungal species with enzymes, enzyme substrates, enzyme classifications, enzyme modifications, enzyme related intellectual property, enzyme retail and applications. The ontology, developed in the OWL language, is the result of the integration of numerous biological database schemas, web accessible text resources and components of existing ontologies. We assess the quantity of implicit knowledge in the Fungal Web ontology by analyzing the range of tags in the OWL files and along with other description logic (DL) computable metrics of the ontology, contrast it with other publicly available bio-ontologies. Thereafter we demonstrate how the FungalWeb Ontology supports its broad remit required in fungal biotechnology by (i) suggesting semantic queries typical of a fungal enzymologist involved in product development, (ii) presenting application scenarios, and (iii) presenting the conceptualizations of the ontological frame able to support these scenarios. Recognizing the complexity of the ontology query process for the non-technical manager we introduce a simplified query tool, Ontologent Interative Query (OntoIQ) that allows the user to browse and build queries from a selection of query patterns. The OntoIQ interface supports users not familiar with writing DL syntax allowing them access to the ontology with expressive description logic and automated reasoning tools. Finally we discuss the challenges encountered during the development of semantic infrastructure for fungal enzyme biotechnologists.

Keywords: Semantic Web, Ontology, Fungi, Enzyme, Description Logic

1 Introduction

Fungi are microorganisms well known for the range of novel enzymes they produce and enzymes of fungal origin are now used in many industrial processes. Their sales contribute strongly to the billions of dollars of revenue made by the biotech industry. The path to product development, namely gene discovery, enzyme characterization, mutational improvement and industrial application is long and fraught with numerous hurdles, both with respect to the domain knowledge and technical challenges. Both within an academic setting and in industrial research and development many decisions are made on incomplete knowledge. This is partly the result of the information overload scientists are currently experiencing and partly since the required knowledge is distributed between numerous disciplines. The current need in the enzyme research and development environment is to have an integrated framework for discovery and decision support, namely an intelligent assistant [1]. This must integrate data from laboratory research, data accessed from distributed database and textual resources such as enzyme product literature, as well as the results of bioinformatics computation. Existing technologies can be assigned to this need for data integration. Specifically a formal knowledge framework is appropriate such as ontologies, supported by tools such as multi agent systems for data retrieval and text mining. Subsequently, advanced query tools with simple semantic query access will enable the fungal enzymologist to address a range of complex questions. To provide a reliable semantic resource in a contemporary research and development environment the scientific and technical span of the ontology must encapsulate a broader and more interdisciplinary range of concepts than is currently accessible in domain specific ontologies. The full range of conceptualizations required by the fungal enzymologist includes fungal species taxonomy, gene discovery, protein family classification, enzyme characterization, enzyme improvement, enzyme production, enzyme substrates, enzyme performance benchmarking, enzyme assays, and market niche. Inclusion of such concepts and instance data related to these subject categories will provide knowledge repositories that support the manager in a scientific discovery process at various times from its inception, through the data production to the result interpretation phase. Development of such a repository is one of the goals of the Fungal Web data integration initiative. The ranges of topics a fungal enzyme biotechnologist must consider are described in Figure 1. Topic boxes overlap according to their co-dependence and are arranged to indicate their sequence in the iterative enzyme development process.

Fig. 1: The biological domains within the scope of the Fungal Web Ontology



2 The FungalWeb Ontology

The FungalWeb Ontology [2] is the result of integrating numerous biological database schemas, web accessible textual resources and interviews with domain experts and the reuse of some existing bio-ontologies. The ontology includes both hierarchical structures supporting full subsumption taxonomies and a broader conceptual frame with novel relationships for specific domain knowledge. The resources for fungus and enzyme related terminologies and concepts come from the resources listed below. In the rest of this paper we refer interchangeably between OWL and DL terminologies.

- NCBI taxonomy database [3]: contains the names of all organisms including fungi.

- NEWT: the taxonomy database maintained by the Swiss-Prot [4].

- BRENDA [5]: a database of enzymes which provides a representative overview of enzyme nomenclature, enzyme features and actual properties.

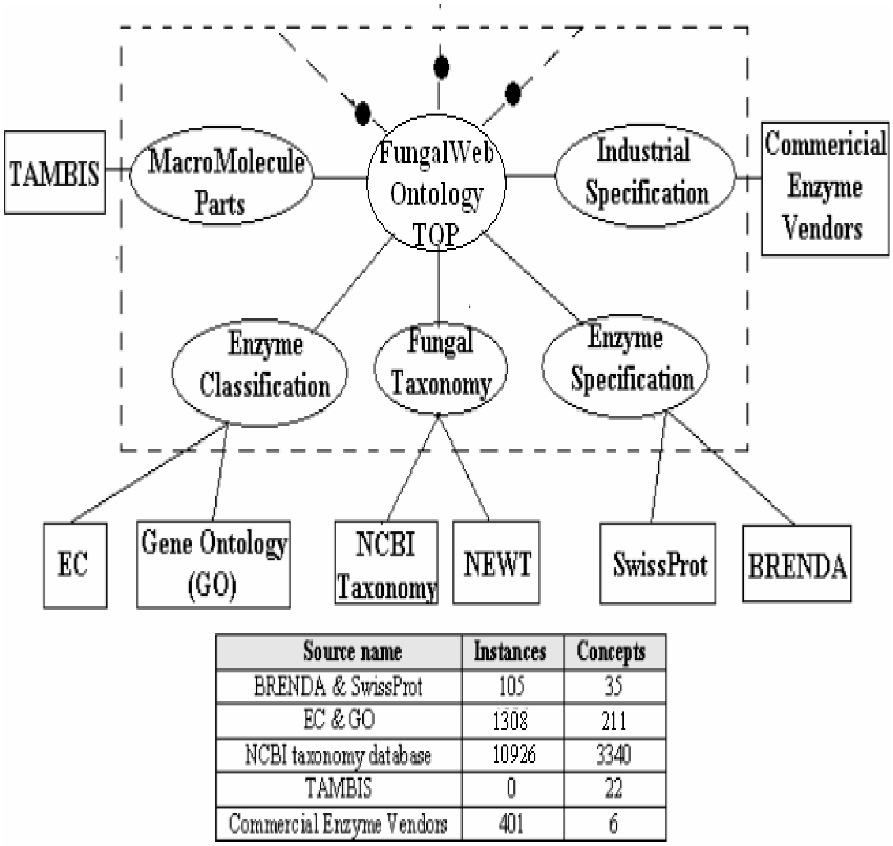
- SwissProt [6]: a protein sequence database providing highly curated annotations, a minimal level of redundancy and a high level of integration with other databases.

- Commercial Enzyme Vendors: Companies that retail enzymes provide detailed descriptions of the properties and benefits of their products on their websites.

- The FungalWeb Ontology also reuses existing domain specific bio-ontologies such as Gene Ontology (GO) [7] and TAMBIS [8].

While efforts to expand the conceptualization are continuing, the ontology currently contains 3667 concepts, 157 properties and 12764 individuals. As shown in Figure 2 most of the terminology in the ontology describes fungal organisms and fungal enzymes. The classification of fungi presented in the ontology is based on phylum, class, order, family, genus and species. Species are considered as the fungal individuals. Enzymes are classified based on catalyzed reactions recommended by the International Union of Biochemistry and Molecular Biology (IUBMB). Enzyme names are considered as the enzyme individuals. Chemical names are considered as individuals of enzyme substrates.

Different properties are defined to relate individuals. For example the property "has been reported to be found in" relates an enzyme individual to a corresponding fungal species individual and is based on the occurrence of a scientific publication citing the existence of an enzyme in a given fungus. Examples of such publications are recorded in the Brenda DB [5]. Figure 3 shows a selection of roles each with corresponding domain and range concepts. Concepts that have children have the number of children included in brackets.



Source name	Instances	Concepts
BRENDA & SwissProt	105	35
EC & GO	1308	211
NCBI taxonomy database	10926	3340
TAMBIS	0	22
Commercial Enzyme Vendors	401	6

Fig. 2. The major resources included within the FungalWeb Ontology.

2.1 Ontology Development

Our goal was to take advantage of the combination of the OWL framework with expressive Description Logics (DL) without losing computational completeness and decidability of reasoning systems. Protégé 2000 [9] was used (with an OWL plug-in) as a knowledge representation editor. The Ontology was designed with a high level of granularity and implemented in the OWL-DL language. The ontology uses and integrated existing bio-ontologies and their conceptualization by merging, mapping and sharing common concepts using logics to build a basic core upon which biotechnology specific concepts have been added. Integration was required on two levels: data integration and semantic integration. Integration of data was achieved by normalizing extracted data formats into a consistent representation.

Domain	**Role**	**Range**
Enzyme (163)	Has_been_reported_to_be_found_in	Fungus
Enzyme	Has_temperature_stability	Temperature_Stability
Enzyme	Has_oxidation_stability	Oxidation_Stability
Enzyme	Has_general_stability	General_Stability
Enzyme	Has_pH_Stability	pH_Stability
Enzyme	Has_storage_stability	Storage_Stability
Enzyme	Has_organic_solvent_stability	Organic_Solvent_Stability
Enzyme	Can_be_renatured	Renatured_Enzyme
Enzyme	Has_purity	Pure Enzyme
Enzyme	Has_specific_activity	Specific_Activity
Enzyme	Has_Ki_Value	Ki_Value
Enzyme	Has_Km_value	Km_Value
Enzyme	Has_inhibitors	Inhibitors
Enzyme	Has_cofactor	Cofactor
Enzyme	Has_turnover_number	Turnover_Number
Enzyme	Has_functional_parameters	TOP
Enzyme	Has_temperature_optimum	Temperature_Optimum
Enzyme	Has_enzyme_specification	TOP
Enzyme	Has_pH_optimum	TOP
Enzyme	Has_temperature_range	Temperature_Range
Enzyme	Has_pH_range	pH_Range
Enzyme	Has_activating_compound	Activating_Compound
Enzyme	Has_EC_number	TOP
Enzyme	Is_able_to_modify	Substrate
Enzyme	Has_activity_towards_substrate	Substrate
Enzyme	Can_be_used_in	Industrial_and_environmental_processes
Enzyme	Can_be_found_in	Commercial_Enzyme_Product
Semantic_word_stem_of_the_substrate_of_the_ enzyme_reaction	Stem_ found_in	Enzyme
Enzyme	Enzyme_name_contains_the_stem	Semantic_word_stem_of_the_substrate_of_the_enzyme_reaction
Enzyme	Encoded_by	Gene
Enzyme	Has_clone	Clone
Enzyme	Has_enzyme_ligand_interaction	TOP
Substrate	Is_activated_by_enzyme	Enzyme
Substrate	Is_a_growth substrate_for	Fungi
Fungi (2110)	Has_been_reported_to_have_enzyme	Enzyme
Fungi	Can_be_used_in	Industrial_and_environmental_processes
Fungi	Has_cell_component	TOP
TOP	Has_Eukaryotes_cell	Fungi_Specification (6)
Fungi	Has_Eukaryotes_cell	Cell (2)
TOP	Has_Fungal_Specification	Fungal_Specification
TOP	Has_Body	TOP
Fungi	Has_pH_Optimum	TOP
Fungus	Grows_on_substrate	Substrate
Commercial_Enzyme_Product	Contains	Enzyme
Commercial_Enzyme_Product	Sold_by	Vendor_Name
Commercial_Enzyme_Product	Has_Industrial_benefits	Industrial_benefits
Commercial_Enzyme_Product	Has_Temperature_Range	Temperature_Range
Commercial_Enzyme_Product	Has_application_description	Application_description
Commercial_Enzyme_Product	Can_be_used_in	Industrial_and_environmental_processes
Application_description	Provides_description_for_product	Commercial_Enzyme_Product
Industrial_and_environmental_processes	Is_using	Commercial_Enzyme_Product
Industrial_and_environmental_processes	Has_enzyme_producer	Vendor_Name
Enzyme_Vendor	Produces_enzyme_used_in	Industrial_and_environmental_processes
Enzyme_Vendor_Name	Sells	Commercial_Enzyme_Product
Industrial_benefit	Benefit_derived_from_using	Commercial_Enzyme_Product
Macro_molecule_part (22)	Part_of	Macro_Molecule
Macro_Molecule (167)	Has_part	Macro_molecule_part
Gene	Regulated_by	Promoter
Promoter	Regulates	Gene
Gene	Encodes	Enzyme

Fig.3. Selected properties, domain and range in the FungalWeb ontology

Semantic integration was achieved by manually identifying relevant data items and reviewing their semantic commonality in order to bring them in a unified frame of reference.

Aptness (considering completeness, consistency and conciseness) of the ontology for its intended application and the scientific integrity was evaluated by posing DL queries with RACER [10], a description logic reasoning system with support for T-Box (concepts) and A-Box (individuals). Syntactic consistency was checked automatically by KR editors and logical consistency was assessed by the DL reasoner. On average, Racer solves the subsumption computations within a fraction of a second. The performance of Racer was dependent on the number of individuals in the ontology. The number of properties did not have an effect on the response time, however we noted the number of property fillers had the strongest influence on the performance with respect to the retrieval of individuals.

Although we sought to validate the biological data and relations by citing their origin (database or literature) or by checking consistency, validation by the domain expert was also necessary. To assist the domain expert in this endeavor we facilitated the use of nRQL [11] to retrieve values of annotation properties used to annotate the ontology resources. These annotations represent metadata (i.e. comments, creator, date, identifier, source name, source URL, version, etc.) but can not be used for reasoning. This capability is additionally useful for the ontology maintenance, versioning and providing proof and trust in a semantic web system.

2.2 Ontology Evaluation

Our ontology was evaluated on the following criteria: occurrences of multiple inheritances, the number of children per concept, depth, and use of OWL tags that store implicit knowledge. Furthermore we compared the FungalWeb ontology to 50 publicly available bio-ontologies [12], in order to assess how amenable it is to knowledge discovery in the context of other ontologies in the same general domain. Our initial analysis involved the simple enumeration of the OWL and RDFS tags in our ontology and 50 other bio-ontologies. The FungalWeb ontology was shown to use the following OWL tags; rdfs:domain 37, rdfs:range 54, rdfs:subProperty 32, rdfs:subclass 3585, owl:Class 5423, owl:disjointWith 4, owl:equivalentClass 1, owl:intersectionOf 2, owl:inverseOf 39, owl:sameAs, 372 owl:ObjectProperty 201 , owl:Restriction 34. In a subsequent analysis all counts pertaining to several *DL interesting* tags (tags which can be considered as DL language elements supporting the derivation of implicit knowledge) [13] were normalised relative to the occurrence of the owl:Class tag and multiplied by 100. To visualise the predominant usage of these tags we discarded scores below 25. This analysis resulted in the patterns or profiles of tags recording implicit knowledge in these bio-ontologies (Figure 4) reflecting their suitability for knowledge discovery using DL reasoning tools. Seven patterns were found, none of which included more than 6 tags. From the frequencies of each pattern we see that pattern 2, containing rdfs:subclass alone, was most prevalent, occurring 29 times. This is not untypical. The same usage pattern was previously identified as the most frequent in previous studies of larger numbers of ontologies randomly selected from the internet and representing a variety of subject domains [12]. The FungalWeb ontology also displayed this pattern indicating that it is not much different from the majority of other bio-ontologies which demonstrate relatively little implicit knowledge. Other well known bio-ontologies exhibited more complex patterns, namely Biopax level 1 [14], pattern 1, and the OWL version of GO (biological process) [7] pattern 5. Pattern 2 was further identified as the pattern most significantly represented with respect to the total numbers of tags in a pattern. The very large GO (biological process) ontology also contributed significant numbers of *DL interesting* tags to pattern 5 which displayed the largest number of tags per pattern within these 50 ontologies.

	Pattern							
Implicit knowledge Tags	0	1	2	3	4	5	6	7
owl:cardinality								
rdfs:domain		*		*				
rdfs:range		*		*				
rdfs:subclass		*	*	*	*	*	*	
rdfs:subProperty					*			
owl:disjointWith		*					*	*
owl:equivalentClass						*		
owl:IntersectionOf						*		
owl:inverseOf					*			
owl:transitiveProperty					*			
owl:oneOf								
owl:symmetricProperty								
owl:functionalProperty								
owl:inverseFunctionalProperty								
Frequency of pattern	7	4	29	2	1	2	4	1
Total tags in pattern	585	1270	83401	218	21	31582	1143	249
Tags per pattern	83	317	2878	109	21	15791	285	249

Fig. 4. Usage patterns implicit knowledge OWL tags in 50 bio-ontologies

The combinations of DL interesting tags found in OWL bio-ontologies were assessed. Numbers of tags in each ontology were normalised relative to the occurrence of the owl:Class tag in the ontology and multiplied by 100. The data was the cleaned to remove insignificant use of tags, namely scores below 25. The FungalWeb ontology exhibited the tag profile found in pattern 2.

We further analysed these ontologies for basic architectural features using DL queries. Using such an approach we rapidly assessed that the FungalWeb ontology has a maximum depth of 12 levels, no occurrences of multiple inheritances and an average of 1 child per concept. It currently contains 3667 concepts, 157 Properties and 12764 individuals. Figure 5 shows the maximum, minimum and average of these parameters in the 50 ontologies studied. These results indicate that the FungalWeb ontology has on average a deeper hierarchy than the selected bio-ontologies. It is clear of dependence on multiple inheritance and has a relatively lean taxonomy falling in the lower range of architectural complexity relative to other bio-ontologies. However it has a higher than average number of concepts and a significantly larger number of individuals. By contrast an early version (biopax-level1.owl) of Biopax [14] exhibited a shallower depth of 6 levels, one occurrence of multiple inheritance, a similar level of children per class, only 27 classes, 121 roles and one instance.

	Depth	Multiple Inheritance	Children per Class	Number of Classes	Number of properties	Number of Individuals
Max.	19	984	2.82	7810	844	3628
Min.	3	0	0.13	3	18	1
Ave.	9	94	0.98	1106	121	341

Fig. 5 DL based evaluation of 50 OWL Bioontologies

Despite the relatively low level of implicit knowledge and moderate architectural complexity in the FungalWeb ontology we illustrate in the next section the high value of large numbers of classes and individuals modeled explicitly in the conceptualization for its intended audience through a number of application scenarios.

3 Application Scenarios

The aim of the following section is to demonstrate the scope of the ontological conceptualization of the FungalWeb Ontology and the relevance of the ontology in supporting a range of cross disciplinary, real world application scenarios. To do this we describe *junction* scenarios where the biotechnologist would seek assistance and interrogate the ontology. We aim to demonstrate the benefit of semantic web technology to the life science manager and accordingly we illustrate how the conceptualization facilitates the tasks listed below. We also discuss the extension of the ontology to house individuals of protein mutation information produced by natural language processing.

- Identification of enzymes acting on polymeric substrates
- Identification of enzyme reaction mechanisms
- Identification of enzyme provenance and common taxonomic lineage
- Identification of commercial enzyme products for enzyme benchmark testing

A scientific narrative illustrating the context and the conceptual frame supporting each of these semantic queries is provided in the following sections. In Section 3.5 queries corresponding to the following scenarios are written in first order predicate logic.

3.1 Ontology for identifying enzymes acting on substrates

Enzyme technologies pose significant opportunities to treat waste products that are a considered nuisance within an industrial process or pollutants in natural environment e.g. for bioremediation. Enzyme technologies can additionally provide access to otherwise wasted natural resources such as cellulose for the production of bioethanol [15]. Within an academic or an industrial enterprise a research manager can frequently poses the question, “Could an enzyme be used to degrade this novel chemical substrate?” To determine if an enzyme is suitable to act on such a substrate a domain expert knowledgeable of enzyme biochemistry would typically evaluate a chemical analysis of the substrate and consider what enzymes carry out modification of the bonds in the substrate. Presented here is an approach where key information in a report authored by an analytical chemist describing the chemical nature of the substrate is used to query the ontology. In this scenario the analytical chemist has analyzed and identified an unknown chemical to be the naturally occurring polymer ‘pectin’ known for its gelling properties and use in food conservation. It is composed of multiple units of the monomer galacturonic acid. The chemist’s description ‘A polymer containing repeated units of galacturonic acid’ (Figure 6A) contains the key semantic terms polymer and galacturonic.

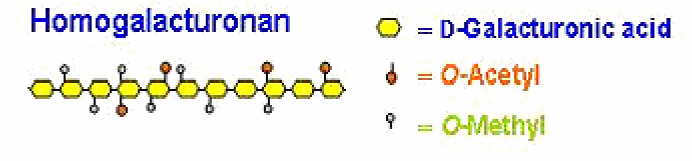


Fig. 6A The structure of poly galacturonic acid, pectin.

Enzyme Class:	EC 3.2.1.67
Enzyme Name:	Exopoly**galacturon**ase: acts on non methoxylated poly**galacturon**ic acid
Common name:	poly**galacturon**ase
Reaction Type:	hydrolysis of O-glycosyl bond, Random hydrolysis of 1,4-?-D-galactosiduronic linkages in pectate and other **galacturon**ans
Other name(s):	pectin depolymerase; pectinase; endopoly**galacturon**ase; pectolase; pectin hydrolase; pectin poly**galacturon**ase; endo-poly**galacturon**ase; poly-?-1,4-**galacturon**ide glycanohydrolase; endo**galacturon**ase; endo-D-**galacturon**ase
Systematic name:	poly (1, 4-?-D-**galacturon**ide) glycanohydrolase

Fig. 6B. Semantically rich enzyme descriptions provided by the systematic classification system introduced by the International Union of Biochemistry (IUB) Enzyme Commission.

In order to facilitate the identification of enzymes able to modify the polymeric substrate pectin it is necessary for the ontology to contain a concept linking enzyme names with a semantic description of the substrates acted on by that enzyme. In the FungalWeb ontology this concept is ‘semantic_word_stem_of_the_substrate_of_the_enzyme_reaction’. Individuals of this concept are produced by a natural language processing tool [16] which summarizes the IUB descriptions of systematic enzyme nomenclature based on reaction mechanisms into nGrams. nGrams represent the most frequently occurring word stems in the IUB descriptions (Figure6B). Since these verbose descriptions often refer multiple times to enzymes by the systematic chemical substrate nomenclature or include specific references to the verbose chemical names, nGrams can act as summaries and individuals of these common-word stems. For example in the case of enzymes such as

Exopoly**galacturon**ase or poly (1, 4-?-D-**galacturon**ide) glycanohydrolase which acts on non methoxylated poly**galacturon**ic acid (a form of pectin) the verbose description can be reduced to the word stem ***galacturon***. This nGram can be instantiated to the concept *Semantic word stem of the enzyme reaction* and have a property relation with individuals of the enzyme to which they relate to, namely a pectinase. Inversely individuals of pectinase enzymes would have property relations with the individuals of the semantic word stems of the substrate, namely poly and **galacturon.** The fungal enzymologist asking which types of enzymes act on pectin would, after first translating the query to relevant classes and properties (Section 3.5 Scenario 3.1 Q1), discover that 7 such enzymes exist. The following pectinases pectinesterase, pectin acetylesterase, endopectinase, exopolygalacturonase, pectate lyase, pectate disaccharide-lyase (exo-polygalacturonate lyase), pectin lyase have the semantic word stem concept instantiated with 'galacturon' and are found when querying the 'word stem' concept using nRQL/DL with Racer. The conceptualization supporting this scenario is shown in Figure 7.

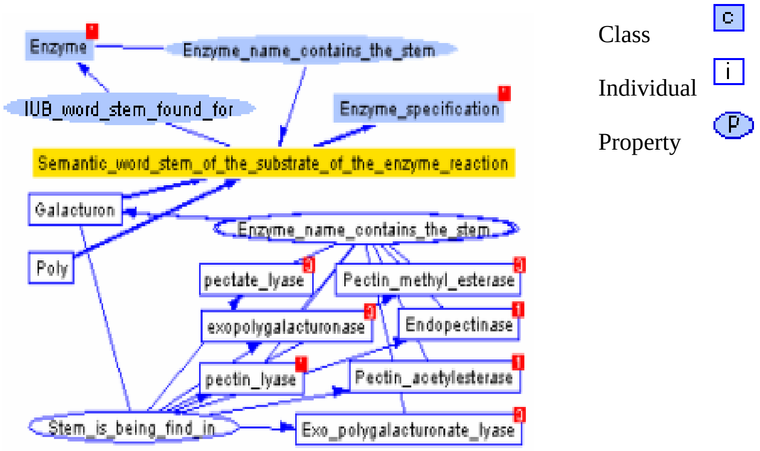


Fig. 7. The conceptual frame supporting the identification of enzymes acting on poly galacturonic. nGrams of IUB reaction mechanism annotations are individuals of the concept 'Semantic word stem of the substrate of the enzyme reaction' highlighted in yellow. Figure 7 was prepared using the ontology visualization tool Growl [17].

Since enzymes in the FungalWeb ontology are conceptualized with respect to their enzyme reaction mechanism we can query the enzyme reaction mechanisms of pectinases and identify them as lyases and two distinct subgroups of hydrolases, esterases and glycosylases. This information is not necessarily explicit from the names of the enzymes. This can be achieved through further use of the subsumption hierarchies and Racer commands to facilitate the selection of all ancestor concepts of the pectinase enzymes identified. This is illustrated in Section 3.5 (Scenario 3.1 Q2). Figure 8 shows the hierarchy in the ontology describing reaction mechanisms of pectinases. Having identified what types enzymes could be used to modify the substrate of interest, the fungal enzymologist would further pose the question 'where are such enzymes found'. In the next section we describe how the FungalWeb ontology can be used to answer this question.

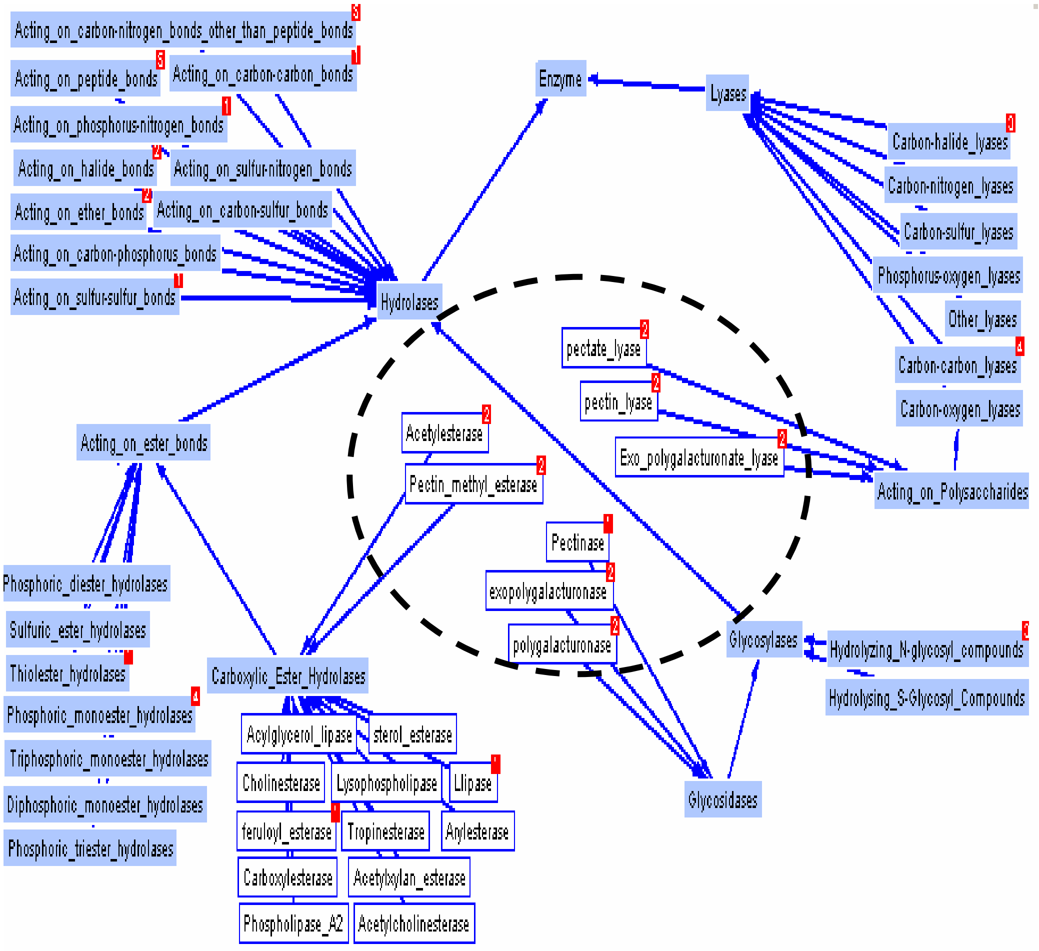


Fig. 8. Conceptual frame supporting the identification of pectinase reaction mechanisms. All decedents of the concept 'Enzyme' relevant to the illustration of the subset of enzymes that are pectinases are displayed. Pectinases are surrounded by the dashed oval. The classes 'Lyases' and 'Hydrolase' are direct child classes of 'Enzyme' that characterize the reaction mechanisms used by pectinases.

3.2 Ontology for determining enzyme taxonomic provenance

Taxonomic provenance (which enzymes are found in which fungal species) of enzymes with industrial potential has long been an interest of microbial biotechnologists. In spite of the recent trend, for finding new genes in environmental DNA samples from diverse environments and cloning genes directly from these sources, scientists continue to use bioinformatics techniques to infer a taxonomic origin for these sequences since such information can provide further highly relevant contextual insights [18]. Taxonomic groups that have provided the most useful enzymes or natural products to date continue to be investigated as providers of new enzymes / genetic material or small molecules [19,20] and the identification of useful enzymes from novel taxa can open up new avenues of discovery [21]. Furthermore laboratory isolation to the taxa of interest can be enhanced on the basis of the knowledge of specific nutritional and growth requirements of the specific taxonomic groups in question. Identifying taxonomic provenance is an important component within the Gene Discovery process.

Having identified enzymes able to modify pectin the same fungal enzymologist wishes to know which fungi are known to produce pectinases and the common lineage that these species share. Identifying the common lineage requires identifying the highest taxonomic group that unites all species known to produce the enzyme of interest, akin to finding a common ancestor. For a small number of species this is a relatively simple task that can be accomplished using online web site resources [22] but it becomes significantly more challenging for a wide number of species producing the same enzyme. Within the FungalWeb Ontology a fungal taxonomy is represented in a deep hierarchy of taxonomic units / concepts. A key property between fungi and enzyme '*has_been_reported to_be_found_in'* permits the identification of species found to have pectinases. Querying this property we can retrieve all fungal individuals in the knowledgebase that are known to produce pectin lyase. Subsequently we can compute the lineage for each species known to produce this enzyme using the Racer. The command *instance types* retrieves the concepts which instantiate each fungal species individual and allows us to find the common lineage between them. Both these queries are illustrated in Section 3.5 (Scenario 3.2). The common lineage of pectin lyase producing fungi, according to the available individuals in the ontology, is the sub phylum of the *Ascomycota* called *Pezizomycotina* known anatomically for producing mycelia that make ascocarps (ascus-bearing structures also called ascomata) with hymenia. Similar types of discovery are cited in the litereature where assertions such as 'large-subunit monofunctional catalase enzymes are, so far, restricted to several representatives of fungi within the *Ascomycota* subphylum *Pezizomycotina'* [23] are found. In the future such identifications could possibly use semantic web technology.

3.3 Ontology for enzyme benchmark testing

When a Gene Discovery platform yields a new enzyme product suitable for an industrial application, be it by directed evolution or more traditional approaches, biotechnologists carry out benchmark performance testing with commercially available competitor enzymes [18]. Standardized assays and conditions are used to determine the likely potency of the new product and to consider whether further enzyme improvement is necessary. Identification of vendors supplying competitor enzymes suitable for a given application is a necessary yet time consuming step in this process.

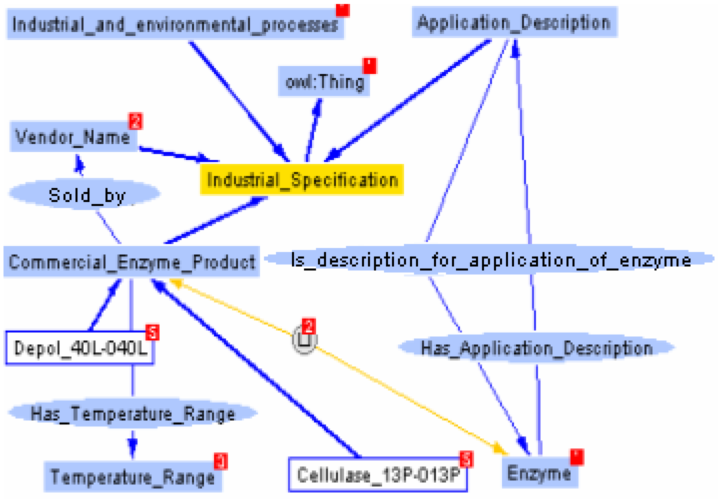


Fig. 9 Conceptual frame supporting the identification of pectinase vendors, the characteristics and application domains of their products.

To benchmark a promising polygalacturonic acid degrading enzyme it is necessary to retrieve all commercial enzyme products sold to an industry using pectinases e.g. the fruit processing industry. We manually instantiated 400 individuals to the industrial specification and are developing agent systems to retrieve these individuals from distributed websites describing product literature in heterogeneous formats. This permits us to query the ontology for enzyme products containing pectinases. Figure 9 shows, Depol_40L-040L and Cellulase_13P-013P are commercial enzyme products that contain pectinases. This is illustrated in Section 3.5 (Scenario 3.3). Using up to date information of this kind the scientist can coordinate a comprehensive benchmarking study of competitor enzymes, flag promising enzymes for further mutational improvement and identify a suitable market niche for new enzymes.

3.4 Ontology for enzyme improvement

Where existing enzymes are known to have a suitable catalytic capability for a given application, and where further benchmarking studies show sufficient enzymatic potency for a given application, there may still be further considerations. Such an enzyme may not have the required functional properties such as pH range or temperature optimum. The enzyme however can still be a good candidate for further mutational enhancement. When considering the improvement of the properties of an existing enzyme there are a range of options [24]. To determine which would be the most successful strategy the scientist would review literature describing the methods employed with close attention to the success of the strategy in improving enzyme properties. Of particular importance is the extent to which a particular approach has improved a property and how much additional improvement is possible. The scientist may ask ‘what is the largest increase in temperature stability of an enzyme achieved by mutational improvement’. Such questions are inherently difficult to answer and require considerable literature review across mutational studies in many different protein families. Ontological and text mining technologies can render and provide access to knowledge concerning the mutational approaches and improvements along with wild type properties of the individual enzymes investigated. The current access route to this information requires manual browsing of distributed database resources such as BRENDA [5] and the scientific literature [3].

```
<Document>
<DocId>PMID: 11377763</DocId>
<Proteins>
<Protein>
<Name>TRX II</Name>
<Organisms>
<Name>T.richoderma reesei</Name>
</Organisms>
<Mutations>
<Mutation>
<Mark>C2-C28</Mark>
<Context>Introduction of multiple arginines into the "Ser/Thr surface" (16), addition of a disulfide C2C28 in the N-terminal
         region (14), and a combination of the disulfide C110-C154 that anchors the R-helix to the adjacent loop region
         together with other weakly stabilizing mutations (15) have considerably increased the thermal stability of TRX
         II.</Context>
 </Mutation>
 </Mutations>
 </Protein>
```

Fig.10 Instance data produced by Mutation Miner [26].

To provide access to such knowledge multiple instance data has been generated using a custom designed tool for natural language processing of scientific full text articles [25]. The instance data shown in the XML format (Figure 10), describes the protein name, the wild type organism, the PMID: PubMed identification number of the paper citing the mutation, the GI: Genbank accession number of the protein, the mutation (Wild Type Residue-Location-Mutant Residue) and the impact of the mutation. This type of instance data could be automatically instantiated to the concepts in the FungalWeb Ontology. The instance data in Figure 10 would then enable our fungal enzymologist to query the ontology for xylanase enzymes that have been modified by site directed mutagenesis. To obtain a deeper semantic access to the impact of a mutation from the <Conext> field, additional concepts are needed within the ontology. Currently we are considering to introduce concepts to permit ranking of such sentences based on descriptions of changes in enzyme properties (shift, increase, more active, fold, destabilize, decrease, remain same), the direction of the change (positive / negative), units of measurement (half life (s), Kcat, hydrolysis efficiency, pH) and the biological property of the enzyme that has been altered (denaturation, catalysis, stability, folding). Appropriate synonym lists for these concepts are additionally being developed. These additions to the FungalWeb Ontology will facilitate semantic access in a manner far superior to manual browsing of texts and database content. Answers to questions such as ‘Find the locations of all mutations in all xylanases that have been reported to have delivered enhanced temperature or pH profile' could soon be enabled. Such a query is currently not answerable from any biology database.

3.5 Application Scenario Queries

This section details the queries to the ontology described in each of the application scenarios and presents them in First Order Predicate logic in place of nRQL syntax [11]. Answers to the queries are included along with relevant racer syntax not describable in logic.

Scenario 3.1: This query finds enzymes where their IUB enzyme_name stems have individuals of 'galacturon'

∃ X :Enzyme (X) /\ Enzyme_name_contains_the_stem (X, Galacturon)

```
Racer returns:    <<<?X :http://a.com/ontology#exopolygalacturonase:>>
                  <<?X :http://a.com/ontology#pectin_lyase:>>
                  <<?X: http://a.com/ontology#Pectin_methyl_esterase:>>
                  <<?X:http://a.com/ontology#Exo-polygalacturonate_lyase:>>
                  <<?X :http://a.com/ontology#Endopectinase:>>
                  <<?X :http://a.com/ontology#pectate_lyase:>>
                  <<?X :http://a.com/ontology#Pectin_acetylesterase:>>>
```

Scenario 3.1: The query retrieves for all enzymes acting on galacturonic acid, the ancestor preceding the common ancestor and used Racer syntax for ancestor concepts 'Instance types'

```
Racer returns:    <:http://a.com/ontology#Lyase:>
                  <:http://a.com/ontology#Hydrolase:>
```

Secnario 3.2: This query finds fungi that have been reported to have a pectin lyase enzyme

∃ X : Fungi (X) /\ have_been_reported_to_have_enzyme (X, Pectin lyase)

```
Racer returns:    <<<?X :http://a.com/ontology#Aspergillus_niger:>>
                  <<?X :http://a.com/ontology#Aspergillus_oryzae:>>
                  <<?X :http://a.com/ontology#Aspergillus sojae:>>
                  <<?X :http://a.com/ontology#Aspergillus_japonicus:>>
                  <<?X :http://a.com/ontology#Glomerella_lindemuthiana:>>
                  <<?X :http://a.com/ontology#Penicillium_expansum:>>
                  <<?X :http://a.com/ontology#Fusarium_oxysporum:>>
                  <<?X :http://a.com/ontology#Penicillium_italicum:>>
                  <<?X :http://a.com/ontology#Penicillium_paxilli:>>>
```

Scenario 3.2 This query retrieves for all fungi that contain pectin lyase the ancestors and uses Racer syntax for 'Instance types'. The common subset thereof is then the desired result

```
Racer returns:
                  <<:?X :http://a.com/ontology#Fungi:>
                  <<:?X :http://a.com/ontology#Ascomycota:>
                  <<:?X :http://a.com/ontology#Pezizomycotina:>
```

Scenario 3.3 The query retrieves the individuals of enzyme products that can be used in fruit processing sold by the vendor *biocatalysts* and are active in the temperature range 50-70 ^{o}Cs

∃X: Commercial_Enzyme_Product (X) /\ Can_be_used_in (X, fruit_and_Vegetable_Processing) /\ Sold_by (X, Biocatalysts) /\ Has_Temperature_Range (X, C50-70)

Racer returns: <<<:http://a.com/ontology#Cellulase_13p-OL3P:>>>

Scenario 3.3 This query retrieves the individuals of vendor name for vendors that sell products containing xylanase enzymes.

∃X: Vendor_name (X) /\ Sells (X, Y) /\ Contains (Y, Xylanase)

Racer returns: <<<:http://a.com/ontology#Dyadic>>>

Fig.11. Queries corresponding to application scenarios

4 Semantic Querying

Beyond the development of the ontology for its intended use we have been forced to recognize the importance of providing simple query access to an ontological conceptualization for non-technically minded experts who acted as domain content evaluators of the ontology. The lisp based syntaxes of DL-based query languages (nRQL and RACER) are difficult for domain experts [27] particularly when seeking to evaluate the ontology for use in their own contexts. For this reason we sought to develop an interactive query tool called the Ontologent Interactive Query tool (OntoIQ) which mirrors the basic query functionalities provided by nRQL used with Racer with the advantage of providing browse and click functionality. Using this tool we were able to give fungal biologists the opportunity to interface with the FungalWeb conceptualization and interrogate it to assess its quality and to allow them to derive insights made possible because of the use of semantic web technology.

4.1 Ontologent Interactive Query

The Ontologent Interactive Query Tool provides a well-organized user interface for all types of users, from the beginner to the professional. Since the current languages for querying the ontology such as nRQL or racer syntax are designed for professional users, the beginner can find it difficult to handle this syntax. The tool introduces much simplified query access based on the employment of a query pattern. OntoIQ is freely available and can be downloaded from the FungalWeb website. http://www.cs.concordia.ca/FungalWeb/OntoIQ.html.

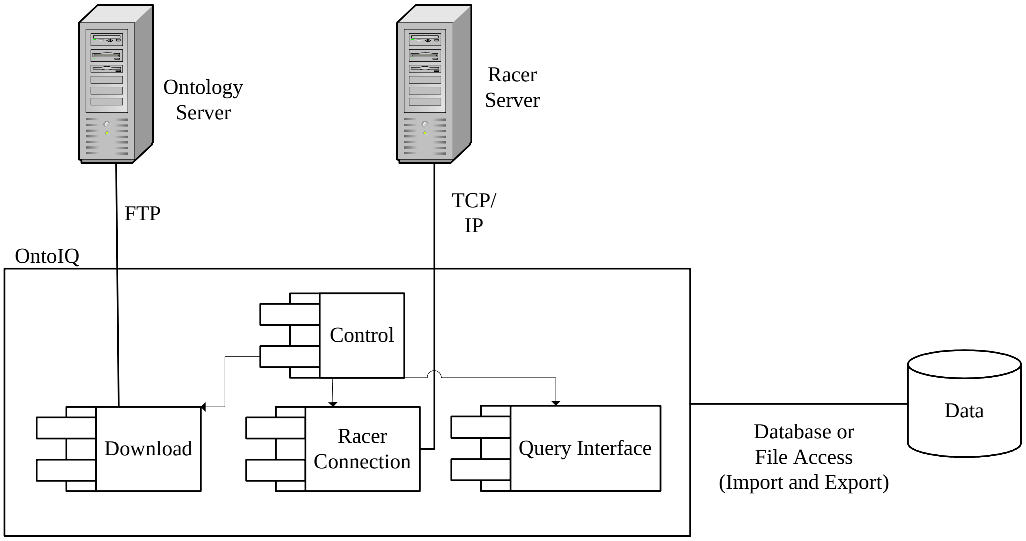


Fig.12. System Architecture of the Ontologent Interactive Query tool (OntoIQ)

4.2 System Architecture

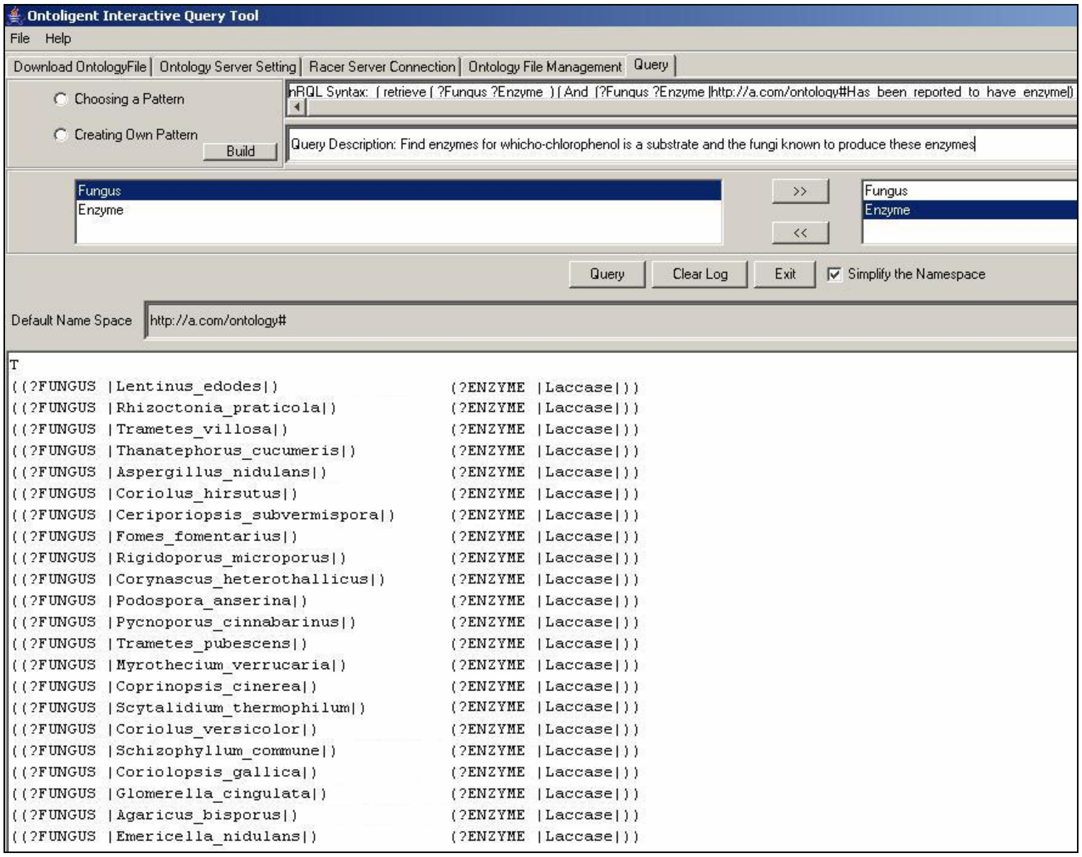
We developed OntoIQ (Figure 12) using a multi-tiered architecture which comprises of the following; an Ontology Server, a Racer Server, an Ontology File Management, and a Query Processor. The Ontology Server allows for specification of server settings for the download of knowledge representation resources hosted locally on a user's machine or available elsewhere on the World-Wide Web. The Reasoning Server allows specification of a remote connection to the Racer Server or to a local executable file of the reasoning engine. The Ontology File Manager serves to load local ontology files to the Racer Server, to permit the checking of file consistency and classify the ontology. The Query Processor provides the gateway to a series of query options. Users are required to select patterns and browse the ontology concepts and axioms in order to provide information specifying what to query from the ontology. The query pane also provides functionality for saving queries for reuse at a later time, translating imported nRQL syntax to a query pattern and export of query results. Direct access to the command line of the reasoner is provided for advanced users wishing to pose queries using Racer syntax not supported by OntoIQ.

4.3 OntoIQ Query Patterns

Simple query patterns such as a *concept* pattern, a *role* pattern, and more complex patterns such as *intersect-conjunction*, *union-disjunction* and *and-combination* patterns exist. Using patterns, users can build up complex queries no matter how much DL-scripting experience they have. Figure 13 summarizes the query patterns available in OntoIQ highlighting the capabilities of the query patterns and the information that must be provided by the user. Building queries through browsing requires a user to identify the pattern required for the type of query and select concepts, roles, the individuals of the concepts or the relationships among these individuals within the ontology. For elementary queries a natural language description of the complex query is synthesized to confirm the query built by the user. After the translation of the pattern to nRQL syntax the query is sent to a Racer server pre-loaded with the ontology. Results are returned to the user in a processed form of the Racer output. Queries can be saved and modified subsequently such that new users get the chance to learn more expressive queries used by other researcher. Query results can be exported in a variety of formats for use by other applications making it conceivable to integrate ontology queries into a workflow.

Pattern	Requires	Query Capability
Concept	A concept and a variable	Is there any instance belonging to a concept? What individuals belong to a concept? Is there no instance belonging to a concept? What individuals do not belong to a concept?
Role	A role and the role's domain and range.	What pairs of individuals are related by role X ? Is there any instance related by a role X ? What individuals of concepts are related by role X?.
Intersect Conjunction	Simple concept or role patterns to combine	What pairs of individuals are related by role X and are related to concept Y by role Z
Union Disjunction	Combines simple concept or role patterns	Retrieve individuals that belong to concept A or individuals that belong to concept B
Combo (And)	Combines any types of above patterns.	Identify individuals of concept A that are related by Role B with individual C. For example, find the names of commercial enzyme products that contain xylanase. In this question, "Enzyme" is the concept name. "Xylanase" is an individual. "Contains" is a role name.

Fig. 13 Query patterns used in Ontologent Interactive Query (Onto-IQ)



nRQL Syntax: (retrieve (?Fungus ?Enzyme) (And (?Fungus ?Enzyme |http://a.com/ontology#Has been reported to have enzyme|)

Query Description: Find enzymes for whicho-chlorophenol is a substrate and the fungi known to produce these enzymes

Fungus
Enzyme

Fungus
Enzyme

☑ Simplify the Namespace

Default Name Space: http://a.com/ontology#

```
T
((?FUNGUS |Lentinus_edodes|)                   (?ENZYME |Laccase|))
((?FUNGUS |Rhizoctonia_praticola|)             (?ENZYME |Laccase|))
((?FUNGUS |Trametes_villosa|)                  (?ENZYME |Laccase|))
((?FUNGUS |Thanatephorus_cucumeris|)           (?ENZYME |Laccase|))
((?FUNGUS |Aspergillus_nidulans|)              (?ENZYME |Laccase|))
((?FUNGUS |Coriolus_hirsutus|)                 (?ENZYME |Laccase|))
((?FUNGUS |Ceriporiopsis_subvermispora|)       (?ENZYME |Laccase|))
((?FUNGUS |Fomes_fomentarius|)                 (?ENZYME |Laccase|))
((?FUNGUS |Rigidoporus_microporus|)            (?ENZYME |Laccase|))
((?FUNGUS |Corynascus_heterothallicus|)        (?ENZYME |Laccase|))
((?FUNGUS |Podospora_anserina|)                (?ENZYME |Laccase|))
((?FUNGUS |Pycnoporus_cinnabarinus|)           (?ENZYME |Laccase|))
((?FUNGUS |Trametes_pubescens|)                (?ENZYME |Laccase|))
((?FUNGUS |Myrothecium_verrucaria|)            (?ENZYME |Laccase|))
((?FUNGUS |Coprinopsis_cinerea|)               (?ENZYME |Laccase|))
((?FUNGUS |Scytalidium_thermophilum|)          (?ENZYME |Laccase|))
((?FUNGUS |Coriolus_versicolor|)               (?ENZYME |Laccase|))
((?FUNGUS |Schizophyllum_commune|)             (?ENZYME |Laccase|))
((?FUNGUS |Coriolopsis_gallica|)               (?ENZYME |Laccase|))
((?FUNGUS |Glomerella_cingulata|)              (?ENZYME |Laccase|))
((?FUNGUS |Agaricus_bisporus|)                 (?ENZYME |Laccase|))
((?FUNGUS |Emericella_nidulans|)               (?ENZYME |Laccase|))
```

Fig. 14 Screenshot of Main Query screen of OntoIQ showing results returned by Racer

4.4 User Interaction Overview

To illustrate the interaction of the user with OntoIQ, we a describe a typical workflow for a user. A screenshot of the main query page of OntoIQ (Figure 14) shows results returned by Racer.

1. Download the OntoIQ tool and run it from the local machine.
2. Provide the IP address for the remote Ontology Server and download an ontology file.
3. Connect OntoIQ with Racer server, load an ontology file, check consistency and run classification.
4. Choose a pattern to build the skeleton of a query sentence.
5. Fill out the concrete information required by the query using the ontology browsing features of OntoIQ. Search functions exist that help users easily choose one concept, instance or role.
6. Transfer a pattern into an nRQL query sentence and send it to the Racer server.
7. Export the query syntax and query results in XML format.

5 FungalWeb Challenges

In the process of employing semantic web technology to build ontology and a large knowledgbase in the domain of fungal biotechnology, we had to deal with a variety of different challenges. Some of the major challenges included working with highly heterogeneous and volatile data, the integration of ontologies implemented in different languages, working with different semantic tools and platforms, and the lack of trustable tools. Ideally a team of experts is required that can sufficiently well overlap with one another in translating needs and requirements. As is typical of such projects frequent and dependable access to a domain expert is an important criterion. Such an expert must also be able to suggest at the outset relevant scientific questions that the content and structure of the ontology must support. This focuses the development of the ontology to the end user needs. The breadth of the subject domain required by fungal enzymologists made the development of the conceptualization a challenge in this regard. We have developed a core ontology that we believe is easily extendable and can support relevant research needs of the enzymologists. Further extension of the ontology into cover concepts relevant to intellectual property is desirable.

Recognition of the utility of the conceptualization has been a challenge since access is difficult for domain experts who are not ready to learn new syntax to pose queries. This is why we developed OntoIQ. The tools still requires users to learn a new approach to querying a knowledge resource. While the challenge posed by the syntax of the query language is now alleviated, the user is now faced with two different challenges that will impact on a users ability to access the desired information, (i) How to illustrate boundaries of the ontological conceptualization (what is included and what is not). This is often facilitated by the use of graphical interfaces such as Growl [17]. However visualization alone is not fully acceptable since it is not integrated with query functionality and secondly a machine could not do this either. (ii) How do users map their queries into patterns that can match the conceptualisation's axioms to retrieve the desired results? Users must become familiar with the properties, and their domain and ranges, used in the conceptualisation as well as the capabilities of each of the query patterns. Domain experts have expressed concerns about the shift in thinking that requires selecting a pattern before formulating the content of a query. This sometimes seems counterintuitive for the beginner. However we considered this challenge in the design of the tool and provide the functionality to save query patterns with a natural language description such that they can be reused at a later time by the same or less advanced users. More advanced users are challenged by the need to compose new patterns to answer what may seem simple queries. This can be addressed by developing an additional tier onto the query tool, namely a natural language query interface which can translate natural language into query patterns. Since the form of prose used by the ontology engineer and user can be sufficiently different this can also impact upon the ease of query answer functionality. For example simple axioms such as *Enzyme-Has_been_reported_to_be_found_in-Fungus* or *Substrate- Is_activated_by_enzyme-Enzyme* could be formulated in a number of additional ways depending on the scientific vocabulary of the user. Despite these issues ontology querying is a new paradigm and OntoIQ makes the existing functionality of nRQL to available non technical knowledge worker, albeit with a learning curve. As such it makes an important contribution to the of evolution knowledge discovery technology.

Also of particular relevance to knowledge discovery is the inclusion of implicit knowledge about a domain within ontologies such that reasoning tools can make non-obvious discoveries. It remains a difficult challenge to make use of the full expressivity of OWL and recent studies [12] have indicated that few ontologies make full use of OWL tags. We have produced an ontological frame of predominantly explicit knowledge that provides fungal enzymologists with semantic access to knowledge relevant to their studies. Since we considered the evaluation of our ontology to include its appropriateness to answer pertinent queries, the ontology design and development evolved with particular query capabilities and axioms in mind. A few of these have been illustrated in the application scenarios (Section 3.5). These are relevant biological questions for which we have demonstrated the usefulness of semantic web technology.

Acknowledgements

This work was financed in part through the Genome Quebec funded project *Ontologies, the semantic web and intelligent systems for genomics*

References

[1] Francis A., Devaney M., Ram A. IRIA: The Information Research Intelligent Assistant, International Conference on Artificial Intelligence (ICAI-00), Las Vegas, Nevada.

[2] Shaban-Nejad A., Baker C.J.O., Haarslev V., Butler G. (2005). The FungalWeb Ontology: Semantic Web Challenges in Bioinformatics and Genomics 4th International Semantic Web Conference (ISWC) November 6-10, 2005, Galway, Ireland. *Lecture Notes in Computer Science,* Vol. 3729, pp. 1063-1066.

[3] Wheeler D.L., Chappey C., Lash A.E., Leipe D.D., Madden T.L., Schuler G.D., Tatusova T.A., Rapp B.A. (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research;* 28(1):10-4. http://www.ncbi.nlm.nih.gov/Taxonomy/ Last accessed 6 December 2005.

[4] Phan I.Q.H., Pilbout S.F., Fleischmann W., Bairoch A. (2003). NEWT, a new taxonomy portal, *Nucleic Acids Research*; 31(13): 3822-3823.

[5] Schomburg I., Chang A., Ebeling C., Gremse M., Heldt C., Huhn G., Schomburg D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Research;* 32(Database issue): D431-3.

[6] Bairoch A. (2000). The ENZYME database in 2000. *Nucleic Acids Research;* 28: 304-305.

[7] Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. (2000). Gene ontology: tool for the unification of biology. *Nat Genet;* 25(1): 25-9.

[8] Baker P.G., Brass A., Bechhofer S., Goble C., Paton N., Stevens R. (1998). TAMBIS—Transparent Access to Multiple Bioinformatics Information Sources. *Proc Int Conf Intell Syst Mol Biol;* 6: 25-34.

[9] Noy N.F., Sintek M., Decker S., Crubezy M., Fergerson R.W., & Musen M.A.. (2001). Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems;* 16(2): 60-71.

[10] Volker H., Ralf M. (2001). RACER System Description. *Proceedings of International Joint Conference on Automated Reasoning, IJCAR?2001*, R. Goré, A. Leitsch, T. Nipkow (Eds.), Siena, Italy. Springer-Verlag, Berlin, pp. 701-705.

[11] Wessel M., Möller R. (2005). A High Performance Semantic Web Query Answering Engine. *International Workshop on Description Logics (DL2005)*, Edinburgh, Scotland, UK.

[12] Baker, C.J.O., Warren R.H., Haarslev V., Butler G. Status Quo of Ontologies in the Public Domain, Submitted.

[13] Patel-Schneider P.F., Hayes P., Horrocks I. (2004). OWL Web Ontology Language Semantics and Abstract Syntax, *W3C Recommendation*, http://www.w3.org/TR/owl-semantics/. Last accessed 6 December 2005.

[14] Luciano J.S. (2005). PAX of mind for pathway researchers. *Drug Discovery Today;* 10(13), 937-942.

[15] Sheehan J., Himmel M. (1999). Enzymes, Energy, and the Environment: A Strategic Perspective on the U.S. Department of Energy's Research and Development Activities for Bioethanol. *Biotechnol Prog;* 15(5): 817-827.

[16] Warren, R.H., Butler, G. Approximate string matching with optimised q-gram generation. In preparation.

[17] Krivov S. (2004). GrOWL Advanced browser for OWL ontologies. Part of *SEEK* project. Java/XML/RDF/OWL. Web Page: http://ecoinformatics.uvm.edu/dmaps/growl. Last accessed 6 December 2005.

[18] Solbak A.I., Richardson T.H., McCann R.T., Kline K.A., Bartnek F., Tomlinson G., Tan X., Parra-Gessert L., Frey G.J., Podar M., Luginbuhl P., Gray K.A., Mathur E.J., Robertson D.E., Burk M.J., Hazlewood G.P., Short J.M., Kerovuo J.J. (2004). Discovery of pectin-degrading enzymes and directed evolution of a novel pectat lyase for processing cotton fabric. *J Biol Chem;* 280(10):9431-8.

[19] Bull A.T., Ward A.C., Goodfellow M. (2000). Search and discovery strategies for biotechnology: the paradigm shift. *Microbiol Mol Biol Rev;* 64(3): 573-606.

[20] Kroken S., Glass N.L., Taylor J.W., Yoder O.C., Turgeon B.G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci USA;* 100(26): 15670-5.

[21] Nishitani Y., Sasaki E., Fujisawa T., Osawa R. (2004). Genotypic analyses of lactobacilli with a range of tannase activities isolated from human feces and fermented foods. *Syst Appl Microbiol;* 27(1): 109-17.

[22] Bischoff J., Domrachev M., Federhen S., Hotton C., Leipe D., Soussov V., Sternberg R. Turner S. The NCBI Entrez Taxonomy Homepage. http://www.ncbi.nlm.nih.gov/Taxonomy/. Last accessed 6 December 2005.

[23] Johnson C.H., Klotz M.G., York J.L., Kruft V., McEwen J.E. (2002). Redundancy, phylogeny and differential expression of Histoplasma capsulatum catalases. *Microbiology;* 148: 1129-1142.

[24] Cowan D.A., Arslanoglu A., Burton S.G., Baker G.C., Cameron R.A., Smith J.J., Meyer Q. (2004). Metagenomics, gene discovery and the ideal biocatalyst. *Biochem Soc Trans;* Pt 2: 298-302.28.

[25] Witte R. and Baker C. J. O. (2005) Combining Biological Databases and Text Mining to support New Bioinformatics Applications. *10 th International Workshop on Applications of*

Natural Language to Information Systems. A. Montoyo et al. (Eds.): NLDB 2005, LNCS 3513, pp. 310-321, 2005

[26] Smith B., Ceusters W., Klagges B., Köhler J., Kumar A., Lomax J., Mungall C., Neuhaus F., Rector A.L., Rosse C. (2005). Relations in biomedical ontologies. *Genome Biology;* 6(5).