# An Experiment on Using Temporal Ontologies to Reason about Localization and Transport of Fungal Proteins

Michel Nathan
Department of Computer Science
and Software Engineering
Concordia University
Montreal, Canada
mar_nath@encs.concordia.ca

Gregory Butler
Department of Computer Science
and Software Engineering
Concordia University
Montreal, Canada
gregb@encs.concordia.ca

## ABSTRACT

The same protein has been observed, through methods of direct assay, to localize to various compartments of the cell. Finding the order in which such subcellular localizations take place contributes toward elucidation of protein pathways and protein-protein interactions. On the other hand, incorporation of the hierarchy of cellular compartments organized as a tree allows a more clear reasoning in terms of the paths taken by proteins that localize to multiple overlapping subcellular sites. In this work, we build an ontology to serve as a knowledge repository for localization of fungal proteins to a hierarchy of major subcellular sites and the order in which such localizations take place. We use this ontology to automatically classify fungal proteins as per their localizations or according to their specific characteristics. Finally, we develop a menu-driven user interface to interact with the constructed ontology. Based on a template of application scenarios, user selections are translated into executable queries to be posed to the system.

## Keywords

temporal order, ontology, multi-site localization, fungal protein

## 1. INTRODUCTION

All biological processes have a beginning and an end. They consist of a set of molecular activities that have to take place in a specific order to ensure a desired outcome. Most processes involve numerous proteins in various stages of their activities. We use two examples of such processes to illustrate the importance of protein localization and their temporal order (Figure 1). First, consider *signal transduction* process. This is a process by which signals and stimuli, originating outside the cell, are captured and converted by the cell to an ordered cascade of biochemical reactions within the cell leading to a desired functional modification. In certain cases this functional change (ex: activation or inhibition of an immune response) is critical to the survival of the organism. In other cases, diseases (such as diabetes) are associated with defects in signal transduction pathways [5]. All intracellular receptors are proteins that mediate the signaling. They need to be located in specific subcellular organelles to ensure successful completion of the signaling process. Another process, with a pathway in opposite direction, is *secretion*. Secretion of substances from the cell, and in particular protein secretion, has an undisputable importance in medicine and biology and numerous protocols have been established to identify the secreted proteins for different classes of species. A comprehensive assessment and application of existing methodologies for identifying secreted fungal proteins may be found in [11]. Secretable proteins are synthesized inside the cell, mostly by the ribosomes. Before they are secreted into the extracellular region, the majority of such proteins undergo a set of activities that take place in various organelles in a specific order. First, the proper folding has to take place in the lumen of the ER. The protein should then be transported to the Golgi apparatus where posttranslational modifications and other functionalization may occur. The proteins are then moved, mostly in secretory vesicles, along the cellular cytoskeleton to the cell membrane. Finally, fusion of the vesicles with the cell membrane is a pre-requisite to unloading of the protein in the extracellular region [1]. There is therefore a definite order in which these proteins are transported across various cellular compartments and a deviation from this order can lead to a dramatically different biological outcome including, but not limited to, complete inhibition of secretion [15].

It is highly desirable to understand the time factor in a localization process. However, the absolute time (ex: datetime stamp) cannot be used to represent biological time in a meaningful manner. Most localization experiments (in-silico or otherwise) search for the presence of specific proteins in major compartments of the cell, without any direct reference to the actual time. When formally representing evidence regarding localization of proteins, what is important, therefore, is if they go to any of the major compartments and if they go to more than one such site, in what order they do so.

In this work we propose a simple method to model the time of localization and build an ontology that integrates knowledge of subcellular structures with data reported on subcellular localization of fungal proteins, compounded with synthetic data to represent the order of such localizations.

## 2. ONTOLOGY DEVELOPMENT

This section describes various stages of development of our ontology. These consist of requirements specification, knowledge acquisition, system implementation and evaluation.


*WWW2007*, May 8–12, 2007, Banff, Canada.
.

**Table 1: 15 major subcellular localization sites**

| Cellular Site | GO Id |
|---|---|
| Bud | GO_0005933 |
| Cell Cortex | GO_0005938 |
| Cell Wall | GO_0005618 |
| Cytoplasmic Vesicle | GO_0031410 |
| Cytoskeleton | GO_0005856 |
| Cytosol | GO_0005829 |
| Endosome | GO_0005768 |
| Endoplasmic Reticulum | GO_0005783 |
| Extracellular Region | GO_0005576 |
| Golgi Apparatus | GO_0005794 |
| Mitochondrion | GO_0005739 |
| Nucleus | GO_0005634 |
| Peroxisome | GO_0005777 |
| Ribosome | GO_0005840 |
| Vacuole | GO_0005773 |

## 2.1 Specification

The ontology is intended to serve two main purposes. First, to incorporate and provide access to knowledge on spatial and temporal localization of fungal proteins, and second, to allow automatic classification of fungal proteins according to certain specific characteristics. Containment and disjointness are the only two spatial relations between subcellular sites that we consider. Other spatial aspects such as adjacency, partial overlap, etc are not covered here. It is also important to note that the system does not capture information on how a protein can go from one given subcellular site to another. Nor does it address the absolute dwell time or duration of any given specific protein in any given specific compartment. Moreover, by temporal order we mean the precedence ordering of localization with respect to the same protein. The relative time of localization of distinct proteins is not covered in this work. For example, the system cannot predict if two proteins localize to the same site simultaneously or one before the other. Finally, the characteristics according to which proteins are automatically classified would include i) their localization to major subcellular sites, ii) whether they are localized to a single or multiple major subcellular sites, and iii) certain specific classes of proteins in which they may be categorized (ex: secreted proteins, membrane proteins).

## 2.2 Knowledge acquisition and Data Source

The main source of information used in this work is shown in Table 3. The GO Cellular Component subontology is written in OWL-DL. OWL is the web ontology language that has been recommended by the World Wide Web consortium (W3C) [12]. Description Logics (DL) is a knowledge representation (KR) formalism that represents the knowledge of an application domain by defining appropriate concepts and relations between them. The building blocks of this KR consist of: i) atomic concepts, ii) atomic roles or properties , and iii) individual constants. An atomic concept is a unary predicate that describes a class of objects (ex: book). An atomic role or property is a binary predicate that represents relationships between objects (ex: a book has a title). An individual constant represents a real word object [3]. OWL-DL is highly expressive (capable of expressing both universal and existential properties) yet decidable (all computations will finish in finite time) [14].

**Table 2: Examples of Competency Questions**

| |
|---|
| Which major sites and in what order does a given protein $p_i$ localize to? |
| Is a given major site $s_j$ the last site that a given protein $p_i$ localizes to? |
| Do all the proteins that end up in the major site $s_j$ pass through a given major site $s_k$ ? |
| List all the specific sites contained within a major site $s_j$ that a given protein $p_i$ localizes to? |
| What percentage of the proteins that pass through the major site $s_j$ pass, beforehand, through the major site $s_k$ ? |
| Are the specific sites $s_1$ and $s_2$ spatially disjoint? |
| List all the proteins whose last major site of localization is $s_j$ ? |
| Which proteins localize to 3 or more major sites? |
| List all the proteins that pass through a specific major site of localization but not through another specific major site? |

**Table 3: Data Source**

| Data Type | Source | Release |
|---|---|---|
| Subcellular sites | http://archive.godatabase.org/ latest-termdb OWL version [2] | Nov. 2006 |
| Protein data | www.yeastgenome.org<br>www.sanger.ac.uk/Projects/ Annotations on S_pombe<br>www.candidagenome.org/ | Dec. 2006 |
| Localization data | www.geneontology.org<br>Annotations on<br>Saccharomyces cerevisia<br>Schizosaccharomyces pombe<br>Candida albicans | Feb. 2006 |

## 2.3 Implementation

This section describes the methodology followed to implement the ontology and the user interface built to interact with it.

### *2.3.1 Conceptualization*

The Gene Ontology (GO) [8] is one of the best-annotated controlled vocabulary for gene products and the GO Cellular Component subontology enumerates most of the biologically relevant intracellular compartments containing established, characterized molecular functions upon which nearly all biologist agree. However, to achieve exhaustiveness, GO Cellular Component considers all possible namings used to refer to various cellular compartments. This leads to the construction of a network whereby the same site is found in more than one hierarchy of subcellular compartments. For example *vacuolar lumen* is classified in GO network in three separate paths, one as a subclass of *organelle*, another as a subclass of *cell*, and still a third one as a subclass of *membrane-enclosed-lumen*. In order to address temporal questions on precedence and succession of localizations, we need to unequivocally determine to which specific site a protein localizes in a given phase of its transport. This is

not possible when the subcellular sites are organized as a graph that includes multiple inheritance. For example, consider the case when a protein, $p$, is reported to first localize to a specific site $s$ and that this site, $s$, is spatially contained in two major organelles $s_1$ and $s_2$ as per GO graph. From this information the system cannot infer which of the two major organelles should be considered the first site of localization of the protein $p$. This problem may be solved if we subdivide the cellular compartments into a hierarchy of sites that cover the entirety of the fungal cell. Such a tree organization would be free of cycles, thus allowing us to uniquely assign ordering to localization sites. Table 1 lists 15 major subcellular sites that disjointly subdivide the fungal cell. Any cellular component may then be uniquely mapped to one of these major sites. We consider the localization to any of the sites specified in Table 1 as an event of interest and propose to model the time ordering of localizations as a series of such events. This constitutes a partial ordering of localizations for each given protein. In addition, the proposed hierarchical structure allows us to investigate spatial localization in subcellular compartments. Here, we can clearly differentiate between proteins that localize to a specific organelle and those that do not.

Following the methodology initially proposed by [7], we begin by developing a list of competency questions that our ontology is expected to answer. These are use-cases that define the requirements of the ontology. In our particular domain of interest, these questions pertain to spatial relations between subcellular compartments, localization of proteins in various compartments as well as the ordering of such localizations when proteins are multiply localized. A set of 50 such questions are formulated, some examples of which are shown in Table 2. These questions serve as guide to define the scope of our ontology and thereby determine its terminology.

The entities in our domain of interest will be represented using classes, attributes and relations. The main concepts of interest in our ontology are fungal proteins and the set of cellular compartments in which they localize. An important decision in ontology design is to determine whether a given concept should be represented as a class or as an instance of a class. In general, concepts that have subclassification should necessarily be represented as classes [4]. Given our intention to consider various classes of proteins and hierarchies of sites, we designate two classes to represent the concepts of *Protein* and *Site*.

We would like our ontology to capture two types of knowledge: i) the spatial taxonomy of subcellular sites, and ii) the proteins that localize to these sites as well as the order of such localizations.

Two main relations that may exist between any 2 subcellular sites are *kind-of* and *located-in* relations. The former indicates that a given site is of the same type as some other site, and the latter pertains to those entities that are spatially located inside another entity and are considered a mereological part of that entity. For example, the vacuolar membrane is a kind of membrane and is also a kind of vacuolar-part and the vacuolar-part itself is located in the vacuole. Hierarchical relation between sites may be described using the notions of ancestry and descendence. The relation *descedent-of* and its inverse, *ancestor-of*, may be used to demonstrate such hierarchical relationships. In our particular domain of interest (subcellular localization), the relation *descendent-of* may be interpreted as meaning any combination of the two relations *kind-of* and *located-in*, i.e. one or both of these relations applied to the concept of *Site*, one or more times. We therefore define *kind-of* and *located-in* relations as sub-properties of *descendent-of* relation. We also define *ancestor-of* as the inverse of *descendent-of* relation.

The temporal order of localization of proteins in subcellular sites may be stated using numerals. Thus, if, for example, a given protein $p_i$ first localizes to the site $s_j$, then to the site $s_k$ then to the site $s_m$ then this fact may be stated using 3 relations: Localizes-To (i, j, 1), Localizes-To (i, k, 2) and Localizes-To (i, m, 3) where the 3 positional parameters refer to an individual protein, an individual site and a temporal order respectively (Figure 2). However, as was mentioned earlier, in Description Logics that forms the basis of our ontology language (OWL-DL), roles are atomic and can be represented only by binary relations between concepts. We therefore have to convert all tertiary relations such as Localizes-To into a set of binary relations before we can represent them as relations in our ontology. In order to achieve this conversion, we propose to define a distinct class for the logical concept of *Localization* and use this class to transform the mentioned tertiary relation into 3 binary relations as shown in Figure 3. To incorporate this knowledge into our ontology, we therefore need to define 3 relations: i) *has-localization*, having the domain *Protein* and the range *Localization*, ii) *has-site*, having the domain *Localization*, and the range *Site*, and iii) *has-ordering*, having the domain *Localization*, and the range *non-negative-integer*. For example, the information Localizes-To (i, j, 1), would be represented using the following 3 binary relations: has-localization($p_i$,$l_k$), has-site($l_k$,$s_j$) and has-ordering($l_k$,1), where $l_k$ stands for the kth instance of localization.

### 2.3.2 Axiomatization

To enhance expressiveness of our ontology, we introduce terminological axioms specific to our domain. These are statements about how various concepts and roles are related to each other [3]. We use equivalent classes to define various classes of proteins. For example, Extracellular-Protein is a protein reported to localize to the extracellular region or any subcellular site that is a part of the extracellular region. This axiom is incorporated into the ontology by the DL statement

$$EP \equiv P \cap \exists hl.((\exists hs.ER) \cup (\forall hs.(\exists do.ER)))$$

where $EP$ and $P$ stand for *Extracellular-Protein* and *Protein* concepts, $ER$ stands for *Extracellular-Region* concept, $hl$, $hs$ and $do$ stand for *has-localization*, *has-site* and *descendent-of* relations. Such equivalent axioms, together with the data on localization instances allow our DL system to make explicit inferences on localization of specific proteins. Namely, upon assertion of a new instance of localization into the ontology, the system automatically verifies if the corresponding protein can be classified as an extracellular protein as per the stated axiom.

We also axiomatized the properties as per usage requirements. We thus made use of nested property as well as other property characteristics such as transitivity and functional relation. For example, the relation *located-in* was defined as a sub-property of *descendent-of* causing all pairs of concepts that are related by the relation *located-in* to become also

related by the relation *descendent-of*. This sub-property axiom thus facilitates taxonomical implications regarding subcellular sites.

### 2.3.3 Integration of localization data

The downloaded GO file contained the 3 sub-ontologies of Cellular Component, Molecular Function and Biological Process. We imported the OWL file into TopBraid Composer, a platform for developing Semantic Web ontologies (http://www.topbraidcomposer.com/). We retained the Cellular Component subontology and stripped the other two. We instantiated the subcellular site classes by generating an instance corresponding to each class and we defined two explicit relations *kind-of* and *located-in* between the generated instances. We also created an instance for each protein found in our downloaded localization data. As temporal ordering information was not availble for proteins known to be present at multiple sites, we assigned precedence ordering to these proteins at random. In the downloaded localization data, each protein is reported to localize to one or more subcellular site. This may be represented by a set of pairs $(p, s)$. We used consecutive numerals (1, 2, etc.) to assign a random ordering to each of the localizations of each given protein in our data set. Thus, for each reported protein we obtained a set of 3-tuples of the form $(p_i, s_j, n)$ indicating that the protein $i$ has subcellular site $j$ as its $n$th reported localization site. We then transformed this set of triplets, in the manner described in section 2.3.1, into three sets of 2-tuples of the form $(p_i, l_k)$, $(l_k, s_j)$, and $(l_k, n)$, where $l_k$ refers to the $k$th instance of localization. We used these three sets and populated our ontology using the relations *has-localization*, *has-site*, and *has-ordering* respectively. Finally, we defined the appropriate relations as per design decisions described in section 2.3.1.

### 2.3.4 User Interface

Users can formulate queries having to do with the spatial and temporal ordering of localization. A template of generic questions was designed based on the competency questions (described in section 2.3.1). We developed a user interface (UI) that builds queries based on User's selection from a set of menus. The main menu allows the selection of the query type (ex: spatial queries, single site localization, multi-site localization, localization ordering, etc). Other menus allow the UI to capture more detailed information to further specify the query. Suppose the user is interested to find out if a given protein passes through a given major site before it passes through a second major site. Through menu selections user identifies the following generic query that would give the desired information:

> Does the protein $p_i$ pass through the major site $s_j$ or one of its descendents before it passes through the major site $s_k$ or one of its descendents?

The UI then prompts the user to choose values from dropdown menus for the parameters $p_i$, $s_j$, and $s_k$. Assuming that the user enters the values *YAL005C*, GO_0005794 (for *Golgi-Apparatus*) and GO_0005777 (for *Peroxisome*) respectively. The UI then generates an executable query that is written in SPARQL (Figure 4). SPARQL is a standardized query language with the appropriate syntax and semantics that allows it to ask and answer queries against RDF graphs [13]. A statement in SPARQL consists of a triplet (Concept,

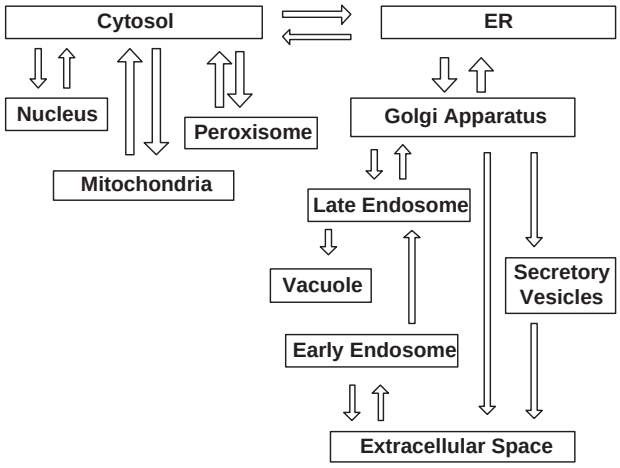


**Figure 1: Example of protein pathways through the cellular organelles**

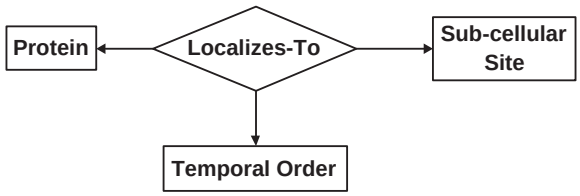


**Figure 2: A tertiary relation between protein, its localization site and the order of its localization**

Relation, Concept) and may contain variables. These variables, as shown in the example in Figure 4, are identified syntactically using words that start with '?'. TopbraidComposer uses the Jena framework to resolve SPARQL queries (http://www.topbraidcomposer.com/community.html). Upon execution of this query, the system returns in sorted order the specific sites contained within the stated major sites that the selected protein localizes to. In this case, *YAL005C* localizes first to GO0005801 (*Golgi-Cis-Face*) and then to GO_0005782 (*Peroxisomal-Matrix*), resulting in a positive answer to the query.

## 2.4 Evaluation

Gomez-Perez [9] advocates evaluation of an ontology during its development and particularly in the conceptualiza-

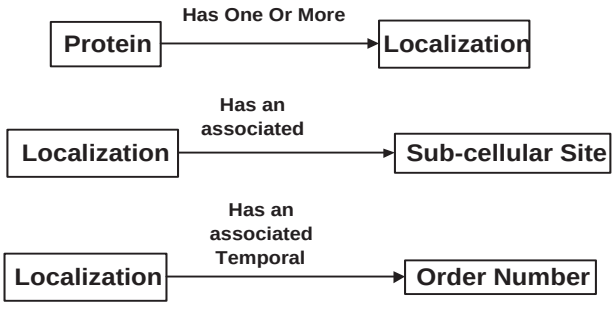


**Figure 3: Reification of tertiary relation into 3 binary relations**

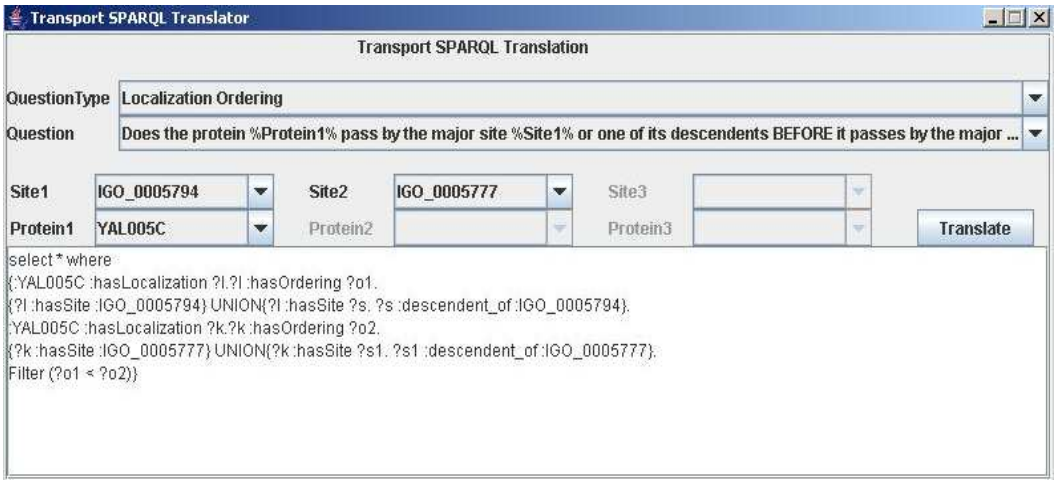


**Figure 4: UI translating a user query into SPARQL**

tion phase before it is propagated into the subsequent phase of implementation. A major aspect of evaluation in large ontologies is consistency. Consistency implies absence of contradiction in individual definitions or what may be inferred from definitions, axioms and other ontologies that are used within the ontology [6]. In our work, we ensured consistency in successive stages of our development. Following each conceptual, relational, functional or axiomatic addition to the ontology, we invoked the class consistency checker tool, built-into TopBraid Composer, to check the semantic validity of the model. TopbraidComposer uses the Pellet reasoner (http://www.topbraidcomposer.com/community.html).

The knowledge base we have developed is based on DL and as such it comprises two components: i) T-Box, containing information about concepts, and ii) A-Box that contains assertions about individuals. Two types of reasoning are performed by the system. T-Box reasoning ensures that all defined relations between concepts hold. This leads to automatic classification of concepts. For example, our system generates a taxonomy of subcellular sites based on the spatial relations defined. Table 4 depicts a small segment of the inferred taxonomy for the site GO_0005798 which corresponds to Golgi-associated vesicle. T-Box reasoning also categorizes various classes of proteins as per superclass - subclass relationships. For example, endosomal, vacuolar, and many other classes of proteins are classified as sub-classes of Cytoplasmic proteins. A-Box reasoning ensures consistency of all defined instances. Various protein classifications of biological interest may thus be achieved through A-Box inferences. In particular, our system automatically classifies all the studied proteins according to the major sites (Table 1) they target. Table 5 depicts a portion of this classification for the sites GO_0005618 and GO_0005634. These ID's correspond to the cell wall and the nucleus respectively. At a broader level, all protein instances were classified to various classes as per their specific characteristics. Examples of the latter classification are membrane proteins and secreted proteins.

It is also proposed that an ontology should be evaluated based on the competency questions that it has to cover [7]. For our system, this type of evaluation is particularly important in cases pertaining to knowledge of temporal localization, as this aspect of our ontology is not validated using T-Box or A-Box inferences. An example of such a query and its result were given in section 2.3.4. In addition to temporal questions, there are other questions of interests that have to be handled through queries. A class of such questions relates to exclusion of localization to certain organelles. For example the user may be interested to find out which proteins among the ones studied pass through the Golgi-Apparatus but not through the peroxisome. For this question the following query is formulated:

Select distinct ?p where
{?p $:has-localization$ ?l.
{?l $:has-site$ :GO_0005794}
UNION {?l $:has-site$ ?s.
?s $:descendent-of$ :GO_0005794}.
?p $:has-localization$ ?k.
?k $:has-site$ ?s1.
Filter (?s1 ! = :GO_0005777).
?s2 $:descendent-of$ ?s1.
Filter (?s2 ! = :GO_0005777)}

The system responds by listing all instances of protein with such a characteristic:

YAL005C
YAL007C
YAL026C
YAR042W
YCL001W
YDR170C
. . .
orf19.1232
orf19.7394
. . . etc.

There are many other specific questions that are best handled using the query system. In an effort to validate the system's capacity to answer such questions, we constructed one or more queries for each generic question producible by the UI. These queries were submitted to the ontology and the results were validated in each case to ensure their correspondence with the reported data.

## 3. DISCUSSION

As per our design decision, 15 subcellular locations were selected to represent the major sites of interest for protein localization (Table 1). We may ask if there are any subcellular sites within a fungal cell that cannot be classified into one of these 15 sites? This can be easily verified for our data set by querying the system:

Select distinct ?p where
{?p :$has-localization$ ?l.?l :$has-site$ ?s.
?s :is-major-site False}

The system's response to this query was an empty set, indicative of the fact that the stated sites do cover the entirety of the fungal cell.

This work may be considered as an experiment that investigates modeling of time in the important question of protein transport. There are various approaches that may be taken to obtain the actual time ordering of localization. Two examples of such approaches are data mining and usage of theoretical knowledge of transport mechanism. Here, we have assumed that localization time ordering is already available and have used synthetic data to validate the working of the developed ontology.

## 4. CONCLUSIONS

We have used a simple modeling of time ordering of localization using basic components available in ontology (class, object property, data types). We have created and used a hierarchy of disjoint sites from the graph of GO for specific usage of protein localization. The OWL file obtained from Gene Ontology has been modified to allow its integration with localization data from fungal protein databases. The developed ontology allows spatial classification of GO sites. It also allows classification of proteins as per their type as well as the major sites in the cell to which they localize. We have used random data to validate that the ontology indeed captures the temporal aspect of localization. We have also developed a user interface to allow non-technical users to generate their desired queries by selecting from a menu made from a template of questions.

## 5. FUTURE WORK

In a previous work [10], we built a predictor for multi-site subcellular localization of fungal proteins. The experimental evidence of localization was obtained from the databases used in the present work (Table 3). We used 178 features covering three types of protein characteristics: i) amino acid compositional features to represent physiochemical and interacting properties, ii) functional motifs to account for family-specific molecular functions, and iii) targeting motifs to serve as indicators of biological pathways in which the protein takes part. One of the outcomes of the mentioned work was the generation of a decision tree depicting the presence or absence of each of the selected features in proteins localizing to each specific subcellular site. This knowledge could be extracted and converted into association lists between localization sites and feature sets. In upcoming work, we intend to expand the present ontology to incorporate: i) terminology for the features potentially implicated in subcellular localization, and ii) a set of rules representing various transport mechanisms that fungal proteins deploy to successively pass through selected organelles to reach their final destination. Feature-localization association sets, such as ones derivable from our decision tree predictor, could then be validated and further investigated using this enahnced ontology in an effort to better understand the relationship between protein features and its localization as well as to improve the predictions.

**Table 4: Segment of subcellular site taxonomy generated by the system**

| Subject | Predicate | Object |
|---|---|---|
| . . . | . . . | . . . |
| GO_0005798 | descendent-of | GO_0044424 |
| GO_0005798 | descendent-of | GO_0043227 |
| GO_0005798 | descendent-of | GO_0044464 |
| GO_0005798 | descendent-of | GO_0005622 |
| GO_0005798 | descendent-of | GO_0031410 |
| GO_0005798 | descendent-of | GO_0005623 |
| GO_0005798 | kind-of | GO_0044422 |
| GO_0005798 | ancestor-of | GO_0030137 |
| GO_0005798 | kind-of | GO_0043227 |
| GO_0005798 | ancestor-of | GO_0030134 |
| GO_0005798 | kind-of | GO_0044444 |
| GO_0005798 | descendent-of | GO_0031998 |
| . . . | . . . | . . . |

**Table 5: Segment of system-generated classification of proteins as per localization to major sites**

| Subcellular site | Protein |
|---|---|
| . . . | . . . |
| GO_0005618 | SPAC14C4.09 |
| GO_0005618 | SPAC17A5.04c |
| GO_0005618 | YLR084C |
| GO_0005618 | orf19.1065 |
| GO_0005618 | orf19.1321 |
| GO_0005618 | orf19.1738 |
| GO_0005618 | orf19.1779 |
| . . . | . . . |
| GO_0005634 | SPAC1002.15c |
| GO_0005634 | SPAC1071.01c |
| . . . | . . . |

## 6. ACKNOWLEDGMENTS

We would like to thank Gopinatha Jakadeesan for his contribution toward the development of the ontology.

## 7. REFERENCES


[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of The Cell, 4th edition*. Garland Science Publishing, New York, NY, USA, 2002.

[2] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature Genet*, 25(1):25–29, 2000.

[3] F. Baader and W. Nutt. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, Cambridge, Massachusetts, USA, 2003.

[4] K. Baclawski and T. Niu. *Ontologies for Bioinformatics*. MIT Press, Cambridge, Massachusetts, USA, 2006.

[5] M. Bjornholm and J. R. Zierath1. Insulin signal transduction in human skeletal muscle: identifying the defects in Type II diabetes. *Biochem Soc Trans.*, 33(2):354–357, 2005.

[6] F. Gomez-Perez. Evaluation of ontology. *International Journal of Intelligent Systems*, 16(3):391–409, 2001.

[7] M. Gruninger and M. S. Fox. Methodology for the design and evaluation of ontologies. 1995. In: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal, Canada, 1995.

[8] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, and R. White. Gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32:D258–D261, 2004.

[9] M. F. Lopez, A. Gomez-Perez, J. P. Sierra, and A. P. Sierra. Building a chemical ontology using methodology and the ontology design environment. *IEEE Intell Syst Appl*, 4(1):37–45, 1999.

[10] M. Nathan. *A Multiple Site Predictor for Subcellular Localization Of Fungal Proteins*. Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, 2006.

[11] N. O'Toole, X. J. Min, G. Butler, R. Storms, and A. Tsang. Sequence-based analysis of fungal secretomes. *Applied Mycology and Biotechnology*, 6:277–296, 2006.

[12] P. F. Patel-Schneider, P. Hayes, and I. Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. 2003. eds. W3C Candidate Recommendation at http://www.w3.org/TR/2003/CR-owl-semantics-20030818/.

[13] E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. 2006. W3C Working Draft 4 October 2006 at http://www.w3.org/TR/2006/WD-rdf-sparql-query-20061004/.

[14] R. Rosati. On the decidability and complexity of integrating ontologies and rules. *Journal of Web Semantics*, 3(1):61–73, 2005.

[15] T. L. Tekirian. The central role of the trans-golgi network as a gateway of the early secretory pathway: Physiologic vs nonphysiologic protein transit. *Experimental Cell Research*, 281(1):9–18, 2002.