

CENTRO UNIVERSITÁRIO FEEVALE

MAGNO SEVERO DE LIMA

EXTRAÇÃO DE SUBCONJUNTOS DE ONTOLOGIAS:
GENEONTOLOGY SLIM

Novo Hamburgo, junho de 2009.

MAGNO SEVERO DE LIMA

EXTRAÇÃO DE SUBCONJUNTOS DE ONTOLOGIAS:
GENEONTOLOGY SLIM

Centro Universitário Feevale
Instituto de Ciências Exatas e Tecnológicas
Curso de Ciência da Computação
Trabalho de Conclusão de Curso

Professor Orientador: Rodrigo Rafael V. Goulart

Novo Hamburgo, junho de 2009.

AGRADECIMENTOS

Gostaria de agradecer a todos os que, de alguma maneira, contribuíram para a realização desse trabalho de conclusão, em especial a minha esposa Cíntia e meu filho Raul pela paciência que tiveram comigo durante o desenvolvimento deste trabalho.

Aos amigos e às pessoas que convivem comigo diariamente, minha gratidão, pelo apoio emocional - nos períodos mais difíceis do trabalho.

Gostaria de agradecer a Prof. Dra. Sabrina Esteves de Matos Almeida e aluna Nicole Pezzi por disponibilizarem seu precioso tempo para ajudarem com informações que foram muito úteis para o trabalho.

Enfim, gostaria de agradecer ao meu orientador pela ajuda na elaboração deste trabalho.

RESUMO

A evolução da Internet e a crescente disponibilidade de informações dificultam a busca de informações co-relacionadas uma vez que os mecanismos de busca atuais não compreendem a semântica das informações. Na biologia esse problema pode ocorrer também, uma vez que em alguma parte do planeta, havendo algum biólogo realizando novos estudos e descobrindo algum novo elemento que foi utilizado para compor um novo produto, essa informação poderia ser útil para outro biólogo que pode estar realizando o mesmo estudo, porém em diferentes partes do planeta. Mas como os mesmos dados são tratados com terminologias diferentes por bases de dados, fica difícil de se encontrar essa pesquisa que muitas vezes acaba levando que a mesma se realize desde o ponto zero. A área a ser estudada trata sobre a geração de ontologias. O projeto realizado visa estudar os passos necessários na extração de ontologias, que tem como objetivo principal gerar ontologias menores a partir de ontologias maiores, para área da biologia. Para esse fim, foram estudadas técnicas utilizadas por ferramentas já existentes no mercado. O propósito deste trabalho é efetuar testes e apresentar os resultados, de um Slim da *Gene Ontology* com foco em uma área de pesquisa desenvolvida na Feevale.

Palavras-chave: Internet, biologia, ontologias.

ABSTRACT

The evolution of the Internet and the increasing availability of information hamper the search for co-related information since the current search engines do not understand the semantics of information. In biology this problem may also occur, since some of the planet, with some new biologist conducting studies and discovering a new element that was used to compose a new product, this information could be useful to another biologist who may be doing the same study, but in different parts of the planet. But as these data are processed with different terminology in databases, it is difficult to find such research often ends up taking it to be held since the zero point. The area being studied is about the generation of ontologies. The project aims to study carried out the necessary steps in the extraction of ontologies, which main goal is to generate ontologies from bigger ontologies more to the biology area. To this end, will study techniques used by existing tools on the market. The purpose of this work is to perform tests and submit the results of a Gene Ontology Slim of focusing on an area of research developed in Feevale.

Key words: Internet, biology, ontologies.

LISTA DE FIGURAS

Figura 1.1 – Ontologias em OWL_____	17.
Figura 1.2 – Código OWL_____	24.
Figura 2.2 – Processo biológico_____	27.
Figura 2.3 – Função molecular_____	28.
Figura 2.4 – Estrutura GO_____	29.
Figura 2.5 – Representação da GO_____	29.
Figura 3.1 – Ferramenta AmiGO_____	31.
Figura 3.2 – Resultado da busca feita na ferramenta AmiGO_____	32.
Figura 3.3 – Resultado da seleção dos itens_____	33.
Figura 3.4 – Tela principal do OboEdit_____	34.
Figura 4.1 – Busca arquivo para análise_____	37.
Figura 4.2 – Selecionar tipo do adaptador_____	38.
Figura 4.3 – Seleção do arquivo na pasta exemplos do OboEdit_____	38.
Figura 4.4 – Inserir o termo para pesquisa_____	39.
Figura 4.5 – Selecionar os termos_____	39.
Figura 4.6 – Mudar cor dos itens_____	40.
Figura 4.7 – Como criar e salvar um filtro_____	40.
Figura 4.8 – Salvar o Slim com o filtro_____	41.
Figura 4.9 – Mostra Slim criado para teste_____	42.

Figura 4.10 – Criando um novo termo	42.
Figura 4.11 – Atribuindo nome ao novo termo	43.
Figura 4.12 – Desfazendo as alterações	44.
Figura 4.13 – Inserindo Base de dados	44.
Figura 5.1 – Material disponibilizado pela especialista	48.
Figura 5.2 – Resultado da busca pelo termo Gag	50.

LISTA DE QUADROS

Quadro 1.1 – Metodologia para ontologias _____ 21.

Quadro 1.2 – Linguagens de representação de ontologias _____ 25.

LISTA DE ABREVIATURAS E SIGLAS

GO	Gene Ontology
SGD	Saccharomyces Genome Database
MGD	Mouse Genome Database
OWL	Web Ontology Language
DAML	Darpa Agent Markup Language
OIL	Ontology Inference Layer
XML	Extensible Markup Language
RDF	Resource Description Framework
XOL	Ontology Exchange Language
SHOE	Simple HTML Ontology Extensions
HTML	Hyper Text Markup Language
RNA	Ácido Ribonucleico
HIV	Human immunodeficiency virus
GAG	Group-specific AntiGen

POL	Polimerase
ENV	Envelope
GPS	Sistema de Posicionamento Global

SUMÁRIO

1 INTRODUÇÃO	13
2 ONTOLOGIA.....	15
2.1 O que são ontologias	15
2.2 Vantagens do uso de ontologias.....	17
2.3 Composição de ontologias.....	18
2.4 Metodologias	19
2.5 Tipos de ontologias	21
2.6 Ferramentas para construção de ontologias.....	22
2.7 Linguagens para ontologias	24
3 GENE ONTOLOGY.....	26
3.1 O que é gene ontology.....	26
3.2 Organização da gene ontology.....	26
3.2.1 Componentes celulares	26
3.2.2 Processos biológicos	27
3.2.3 Função molecular.....	27
3.3 Estrutura da gene ontology	28
3.3.1 Árvore de representação.....	29
3.3.2 Anotação GO.....	29
3.4 Ferramentas GO	30

3.4.1 AmiGO.....	30
3.4.2 OboEdit	33
3.4.3 Anotação no OboEdit.....	34
4 GO SLIM	36
4.1 O que é Go Slim.....	36
4.2 Por que usar Go Slim.....	36
4.3 Produzindo um Slim utilizando o OboEdit.....	36
4.3.1 Produzindo um Slim Passo a passo	37
4.3.2 Editando um Slim	41
4.3.3 Considerações sobre o OboEdit.....	44
5 SLIM PARA PESQUISA EM HIV	45
5.1 Introdução.....	45
5.2 Proposta	46
5.3 Aplicação.....	46
5.4 Material disponibilizado pelo especialista.....	46
5.5 Genes e produtos gene.....	47
5.6 Criação do Slim.....	48
CONSIDERAÇÕES FINAIS.....	52
REFERÊNCIAS BIBLIOGRÁFICAS	54
ANEXO – SLIM HIV GERADO NO OBOEDIT	55

1 INTRODUÇÃO

Pesquisadores da área da Biologia, atualmente, perdem muito tempo e esforço na busca de todas as informações disponíveis sobre cada pequena área de investigação. Fica ainda mais difícil pelas grandes variações de terminologias que podem ser de uso comum, num dado momento, que inibem a eficiência da pesquisa de ambos, computadores e pessoas. Por exemplo, buscar todos os produtos genéticos que estão relacionados a um antibiótico, duas ou mais bases de dados pode tratar uma mesma terminologia de um modo diferente tornando difícil para humanos e computadores encontrar termos equivalentes. Para auxiliar neste processo surgiu o consórcio *Gene Ontology*.

O projeto *Gene Ontology* (GO) é um esforço de colaboração para enfrentar a necessidade de coerência na descrição de produtos genéticos em diferentes bases de dados. O projeto começou com a colaboração entre organismo modelo de três bases de dados, *FlyBase* (*Drosophila*), *Saccharomyces Genome Database* (SGD) e do *Mouse Genome Database* (MGD), em 1998. Desde então, o GO Consórcio tem crescido para reunir bancos de dados, incluindo os maiores repositórios do mundo de vegetais, animais e genomas microbianos (www.geneontology.org GO, 2008).

De forma simples, para elaborar ontologias, definem-se categorias para as coisas que existem em um mesmo domínio. Ontologia é um "catálogo de tipos de coisas" em que se supõe existir um domínio, na perspectiva de uma pessoa que usa uma determinada linguagem (HINZ, 2006 apud SOWA, 2000). Trata-se de "uma teoria que diz respeito a tipos de entidades e, especificamente, a tipos de entidades abstratas que são aceitas em um sistema com uma linguagem" (HINZ, 2006 apud CORAZZON, 2002).

O projeto *Gene Ontology* foi desenvolvido, estruturado e controlado por 3 vocabulários (ontologias) que descrevem, produtos gene em termos dos seus processos biológicos associados, componentes celulares e funções moleculares, em uma espécie de forma independente. Há três aspectos distintos para este esforço: primeiro, o desenvolvimento

e a manutenção das ontologias, em segundo lugar, a anotação de produtos genes, o que implica em fazer associações entre ontologias, os genes e a genética, e o terceiro, o desenvolvimento de ferramentas que facilitem a criação, manutenção e uso de ontologias. O uso de termos GO, com as bases de dados colaborando, facilita as consultas uniformes. Com os vocabulários controlados e estruturados de forma a poderem ser consultados em diferentes níveis: por exemplo, é possível usar GO para encontrar todos os produtos genéticos envolvidos no genoma do rato. Esta estrutura permite também anotar propriedades para atribuir aos produtos gene ou genes em diferentes níveis, em função da profundidade de conhecimento sobre essa entidade.

O foco principal deste estudo é utilizar a GO Slims que extrai um subconjunto da GO. Utilizar toda a GO inclui muitos termos na utilização de inferências. As GO Slims são criadas pelos usuários de acordo com suas necessidades, e podem ser específicas para espécies ou para determinadas áreas da ontologia. A partir deste estudo viabilizar a criação de um Slim para uma área da biologia utilizando a ferramenta OboEdit.

Nos próximos capítulos são abordados temas como Ontologia, sua organização, composição, estrutura, metodologias, algumas linguagens e ferramentas para desenvolvimento e manutenção de ontologias. Também é explicado como funciona o consórcio *Gene Ontology*, sua organização, composição e estrutura. No capítulo 4 é feita uma introdução ao Go Slim e para finalizar, no capítulo 5 é desenvolvida a proposta do trabalho que é a geração de um Slim específico para uma área da Biologia.

2 ONTOLOGIA

Neste capítulo é apresentada definições de ontologia bem como as vantagens de uso e a sua composição, metodologias e algumas ferramentas para construção e linguagens para ontologias.

2.1 O que são ontologias

Ontologias, na filosofia, têm sido usadas para descrever domínios naturais, ou seja, as coisas naturais do mundo como os tipos de existências e as relações temporais. Ontologia é um ramo da filosofia que lida com a natureza e a estrutura do ser. Esse termo foi inserido por Aristóteles em *Metafísica*. Termo também adotado pela comunidade de inteligência artificial para referir-se a conceitos e termos que podem ser usados para apresentar algumas áreas do conhecimento ou construir uma representação deste. Apesar de sua divulgação ao longo do tempo em que vem sendo utilizado, ainda não há um consenso, principalmente na comunidade da Ciência da Computação sobre a semântica do termo “ontologia”.

Uma ontologia define os termos utilizados para descrever e representar uma área de conhecimento. Ontologias são utilizadas pelas pessoas, bases de dados e aplicações que necessitam de partilhar informações de um domínio (W3C, 2004).

Segundo www.geneontology.org GO (2008), ontologias são estruturadas como grafos acíclicos dirigidos, que são similares às hierarquias, mas diferem na medida em que o prazo mais especializado (filho) pode ser associado a mais do que uma expressão menos especializada (pai). Por exemplo, o processo biológico *hexose biosynthetic* tem dois pais, processo *hexose metabólico* e processo monossacarídeo *biosynthetic*. Isto é, processo *biosynthetic* é um tipo de processo metabólico de uma *hexose* que é um tipo de monossacarídeo. Quando qualquer gene *hexose biosynthetic* envolvidos no processo é anotado

para o presente prazo, é automaticamente anotado para ambos os processo hexose metabólica e processo monossacarídeo *biosynthetic*.

Segundo Smith (2007), Ontologia é a ciência do que é, do real e das estruturas de objetos, propriedades, eventos, processos e relações em todos os domínios da realidade.

Segundo Matos (2008) aput Almeida (2003), historicamente o termo ontologia tem origem no grego *ontos*, ser e *logos*, palavra. É um termo introduzido na filosofia com o objetivo de distinguir o estudo do ser humano como tal, do estudo de outros seres das ciências naturais. A origem é a palavra aristotélica “categoria”, que pode ser usada para classificar e caracterizar alguma coisa.

Dentre as várias definições para ontologias a que mais é encontrada freqüentemente na literatura, e que é apresentada por Breitman (2006) proveniente de Gruber em seu artigo “*A Translation Approach to Portable Ontology Specification*”: “*Uma ontologia é uma especificação formal explícita de uma conceitualização compartilhada.*”

É importante explicar o significado das palavras contidas nesta frase. A palavra “conceitualização” traduz um modelo abstrato de algum fenômeno que identifica os conceitos relevantes para o próprio. A palavra “explícita” significa que os tipos de conceitos usados e as limitações do uso destes conceitos devem ser definidos de forma clara. A palavra “formal” refere-se que a ontologia deve ser passível de ser processada por uma máquina. E por fim a palavra “compartilhada” que reflete a noção de que ontologia captura um conhecimento consensual, quer dizer que, esse conhecimento não deve ser restrito a algumas pessoas mas sim aceito por um grupo de pessoas (BREITMAN, 2006).

A Figura 1.1 ilustra uma ontologia construída em uma linguagem OWL.

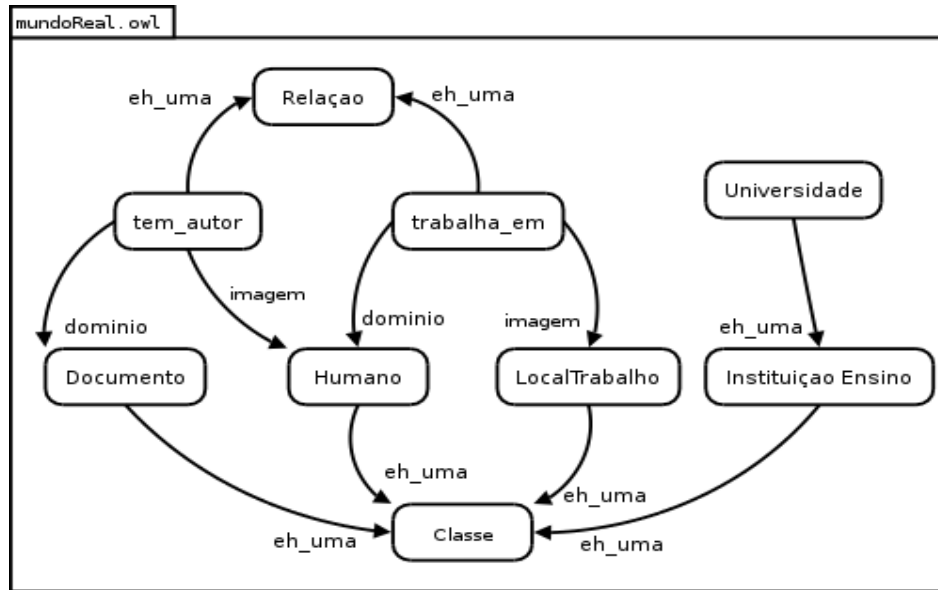


Figura 1.1: Ontologia em OWL

Fonte: <http://www.linux.ime.usp.br/~cef/mac499-04/monografias/rec/ghsilva/mundoReal.png>

2.2 Vantagens do uso de ontologias

Agora que sabemos o que é ontologia, podemos destacar as vantagens em usá-las. Abaixo é apresentada uma lista com as principais vantagens da utilização de ontologias na computação, segundo (HINZ, 2006):

- Fornecem vocabulário para representação do conhecimento.
- Vocabulário que tem por trás uma conceitualização que o mantém. Evitando assim interpretações imprecisas deste vocabulário.
- Permitem compartilhamento de conhecimento, sendo assim, caso exista uma ontologia que modele adequadamente certo domínio de conhecimento, esta pode ser compartilhada e usada por desenvolvedores de aplicações dentro desse domínio. Explicando melhor, considere que exista uma ontologia para o domínio de locadoras. Uma vez que esta ontologia está disponível, muitas locadoras podem montar seus catálogos usando o vocabulário colocado por esta ontologia sem a necessidade de fazer uma nova análise do domínio locadoras.
- Disponibiliza uma descrição exata do conhecimento. Diferente da linguagem natural, em que as palavras podem ter sua semântica muito diferente conforme seu argumento, a ontologia pode ser escrita em linguagem formal, não deixando espaço para o intervalo semântico existente na linguagem natural. Um exemplo

seria quando existe palavra de dupla interpretação como “rede”, a pessoa pode estar falando de uma rede de pescar ou uma rede para descansar. A interpretação da palavra pode ser atribuída a um conceito ou outro dependendo do estado mental da pessoa.

- O mapeamento da linguagem da ontologia é possível ser feito sem que com isso seja mudada a sua conceitualização, ou seja, a mesma conceitualização pode ser publicada em diversas línguas.
- É possível estender o uso da ontologia genérica de forma que ela se adapte a um domínio específico. Exemplo, se um indivíduo precisa de uma ontologia sobre motos para desenvolver uma aplicação é somente necessário encontrar uma ontologia sobre o domínio genérico de veículos, pode usar essa ontologia estendendo para o domínio da aplicação que neste caso é de motos.

Existem muitas outras vantagens em utilizar ontologias, essas que foram apresentadas são as principais, as outras na sua maioria são decorrentes destas. (HINZ, 2006).

2.3 Composição de ontologias

Os componentes básicos de uma ontologia são classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados). (HINZ apud GRUBER, 1996).

Segundo Hinz (2006, p. 25), De forma resumida, uma ontologia é composta de: Domínio, Vocabulário, Glossário, Tesouro, Taxonomia e Rede Semântica.

O domínio define a abrangência da ontologia. Como dito em (HINZ apud RUSSEL; NORVIG, 2004, p. 245), o domínio de uma representação do conhecimento é uma parte do mundo sobre o qual deseja-se expressar algum conhecimento. Este domínio define os limites onde é definida toda a estrutura da ontologia.

Ainda segundo Hinz, vocabulário é composto de classes e instâncias. Classes apresentam conceitos dentro do domínio estimado. Conceitos que representam um conjunto abstrato de objetos.

Instâncias são eventos específicos de objetos em relação à classe avaliada. Uma instância apresenta conceitos, de uma forma individualizada, única e sólida, fazendo referência a um objeto real (documento, páginas da intranet e base de dados).

Dentro do vocabulário, as classes possuem características. Estas características são determinadas pelos seus atributos ou propriedades. Atributos que possuem valor nas instâncias. A taxonomia é composta das relações hierárquicas entre as classes. Essa relação define a estrutura onde classes pai possuem classes filhos, com atributos herdados das classes pai.

A composição das taxonomias se dá através de termos ou conceitos sobre o universo da informação registrada. Assim como o tesouro define conceitos, mas também permite outros tipos de relacionamentos entre os conceitos. Usando taxonomias ou tesouros, o usuário poderá optar por qual categoria de informação ele deseja acessar, combinando com a busca por palavra chave. Esse tipo de busca fornece resultados mais eficientes no acesso à informação. A deficiência apresentada é a não disponibilidade de uma linguagem padrão que descreva sua estrutura. Isso impede o compartilhamento dessa estrutura por diferentes sistemas, à construção de ferramentas de ajuda no desenvolvimento das estruturas e a utilização de sistemas de manipulação de regras de restrição (HINZ, 2006).

2.4 Metodologias

Encontra-se pouco desenvolvido o processo de construção de ontologias. Grande parte dos desenvolvedores usa seus próprios critérios para criar suas ontologias (HINZ, 2006).

Com a falta de uma metodologia comum para o desenvolvimento, torna-se comum entre os desenvolvedores de ontologias passarem diretamente do passo de aquisição de conhecimento para o passo de implementação, o que acarreta nos seguintes problemas:

- Modelos conceituais de ontologia ficam implícitos no código da implementação.
- Dificuldades no reuso da ontologia, o *design* da ontologia e as decisões do projeto estão implícitos no código.
- Ocasionalmente problemas de comunicação devido às dificuldades que o *expert* no domínio da ontologia tem para entender o código da implementação. Torna-se um sério problema, pois ele tende a ser a principal fonte de informação sobre o domínio.

- Causa dificuldades no desenvolvimento de ontologias complexas, pois a passagem da aquisição de conhecimento para a implementação é abrupta.

Para evitar que estes tipos de problemas que foram citados acima ocorram deve-se seguir alguma metodologia. Abaixo é apresentado um quadro com algumas metodologias e suas definições:

Metodologia	Referência	Fases
Enterprise Ontology	(USCHOLD;KING,1995)	<ul style="list-style-type: none"> • Identificar a proposta da ontologia. • Construir a ontologia capturando, codificando e integrando conhecimento apropriado a partir de ontologias existentes. • Avalia a ontologia. • Documenta a ontologia.
TOVE (Toronto Virtual Enterprise)	(GRUNINGER;FOX,1995)	<ul style="list-style-type: none"> • Capturar cenários de motivação. • Formular questões de competência informal. • Terminologia Formal. • Formular questões em FOL. • Especificar axiomas. • Avaliar a ontologia.
Methontology	(GÓMEZ PÉREZ, 1998)	<ul style="list-style-type: none"> • Especificar o requisito. • Conceitualizar o domínio do conhecimento. • Formalizar o modelo conceitual em uma linguagem formal. • Implementar um modelo formal. • Fazer a manutenção de ontologias

		implementadas.
On-To-Knowledge	(STAAB et al.,2001)	<ul style="list-style-type: none"> • Requisitos da ontologia são capturados e especificados. • Questões de competência são identificadas. • Ontologias potencialmente reusadas são estudadas e uma versão “draft” da ontologia é construída.

Quadro 1.1: Metodologias para ontologias

Fonte: <http://ppginf.ucpel.tche.br/TI-arquivos/2006/VerlaniHinz/PPGINF-UCPel-TI-2006-2-10.pdf>

As metodologias apresentadas acima possuem características e abordagens diversas.

Não é possível fazer a união destas propostas apresentadas em uma única metodologia. Para verificar a utilidade das metodologias e compará-las, é necessário avaliar a ontologia resultante da aplicação de cada metodologia.

Segundo Hinz (2006) apud Gómez-Pérez (1999), “Apesar dos princípios de construção propostos, ainda não existem ferramentas de metodologia que guiem uma elaboração completamente padronizada, como em outros tipos de software. Em consequência disso, não há métodos sedimentados de validação, verificação, desenvolvimento e mesmo documentação de ontologias, que muito facilitaria seu reuso e aplicação. Engenharia e metodologias têm sido estudadas e propostas, visando preencher este vácuo”.

2.5 Tipos de ontologias

Segundo Becker (2005) apud Muñoz (2002), há vários critérios adotados para classificar as ontologias existentes. De acordo com Becker (2005) apud Guérios (2005), as ontologias podem ser classificadas quanto ao tipo e quanto à profundidade.

Quanto ao tipo, Becker (2005) apud Guarino (1998) classifica as ontologias nas seguintes categorias:

- **Ontologias Genéricas:** não dependem de um problema ou domínio em particular. Descrevem conceitos gerais, tais como espaço, tempo, matéria e objeto, ou seja, teorias básicas do mundo, de caráter abstrato e que sejam aplicáveis a qualquer

domínio (que constituam conhecimento de senso comum). São construídas com o intuito de serem especializadas na definição de conceitos em uma ontologia de domínio;

- **Ontologias de Domínio:** aplicam-se a um domínio do conhecimento, descrevendo conceituações particulares deste, por exemplo, de Medicina ou Direito. Constituem o tipo mais comumente desenvolvido;
- **Ontologias de Tarefas:** descrevem o vocabulário relacionado a uma atividade ou tarefa genérica, independentemente do domínio desta, tal como diagnose ou vendas;
- **Ontologias de Aplicação:** tal como as Ontologias de Tarefas, elas descrevem conceitos relacionados à execução de uma atividade ou tarefa, contudo, tais conceitos são dependentes do domínio e da tarefa representados;
- **Ontologias de Representação:** explicam as conceituações que fundamentam os formalismos de representação de conhecimento;

2.6 Ferramentas para construção de ontologias

Neste item é explicado o funcionamento de algumas ferramentas utilizadas na construção e edição de ontologias, não será um estudo completo, pois seria impossível apresentar todas as ferramentas disponíveis atualmente.

Como se trata de um trabalho difícil, qualquer ajuda na construção de ontologias pode representar ganhos significativos. Alguns exemplos de ferramentas para construção de ontologias são apresentados a seguir:

OilEd – é um simples editor, funciona como o “NotePad” dos editores de ontologias. Ele oferece suporte a ontologias desenvolvidas nas linguagens DAML+OIL e OWL. Essa versão não apresenta suporte para nenhuma metodologia de desenvolvimento de ontologias. O OilEd não tem suporte para a integração, versionamento e argumentação entre ontologias. Ele simplesmente permite ao usuário escrever ontologias e demonstra como usar o verificador FACT para comparação das mesmas, ou seja, checar se a ontologia desenvolvida está de acordo com o meta-modelo.

OntoEdit – é um editor gráfico para edição de ontologias, permite a inspeção, navegação, codificação e alteração de ontologias. Este editor foi desenvolvido dentro do

projeto On-To-Knowledge. Ele também implementa um processo específico para construção de ontologias. Seu funcionamento ocorre da seguinte maneira: na primeira fase são coletados requisitos que descrevem as atividades em que a ontologia dará suporte. De um modo geral, essa tarefa será realizada pelos especialistas do domínio acompanhados pelos especialistas da modelagem. Essa fase deverá gerar subsídios que conduzirão o engenheiro de ontologia na decisão sobre os conceitos relevantes e a estrutura hierárquica na ontologia. Existe também a fase de refinamento que uma ontologia madura é produzida e orientada a aplicação de acordo com as especificações inseridas na fase anterior. No OntoEdit as ontologias são armazenadas em diferentes bancos relacionais e podem ser desenvolvidas em XML, FLogic, RDF(S) e DAML+OIL. (BREITMAN, 2006).

Ontolíngua – é um conjunto de serviços que permite o desenvolvimento de ontologias compartilhadas entre grupos. Também permite acesso a uma biblioteca de ontologias, tradutores para linguagem e um editor para criar e navegar pela ontologia (HINZ, 2006).

Protegé – ambiente interativo para projeto de ontologias com código aberto, que proporciona uma interface gráfica para edição de ontologias e uma arquitetura para criação de ferramentas baseadas em conhecimento. A arquitetura é modulada e permite a inserção de novos recursos (HINZ, 2006).

Protegé é um ambiente para criação e edição de ontologias, ferramenta de software livre, desenvolvida em Java, tem sua arquitetura aberta, permitindo que os usuários possam desenvolver novos *plugins* capazes de aumentar o conjunto de funcionalidades da ferramenta. O Protege também utiliza um mecanismo de inferência para a verificação de ontologias e a classificação automática. É uma das ferramentas mais utilizadas no desenvolvimento de ontologias em OWL junto com OilEd (BREITMAN, 2006).

Chimaera – é um sistema de software que ajuda o usuário na criação e manutenção de ontologias. Suas principais funções são: combinação de múltiplas ontologias e diagnóstico de ontologias individuais ou múltiplas. Ajuda o usuário na tarefa de carregamento de bases de conhecimento em diversos formatos, reorganização taxionômica, resolução de conflitos com nomes, edição de termos e busca de ontologias. A intenção original desta ferramenta era ser aproveitada para combinar fragmentos de bases de conhecimento heterogêneas.

2.7 Linguagens para ontologias

Muitas linguagens foram criadas para representação de ontologias. Exemplos como: XOL, SHOE, DAML, RDF, RDF(S), OIL e OWL.

Dentre essas linguagens apresentadas, a OWL é a mais promissora e sofisticada do que suas linguagens bases, porque seus conceitos podem ser especificados por combinações lógicas como interseção, junção, ou complemento de outros conceitos. O OWL pode indicar que uma propriedade é transitiva, funcional, transitiva ou inversa, em relação a uma outra propriedade, instâncias pertencem à quais conceitos, que conceitos e propriedades podem ser utilizadas entre indivíduos.

O OWL foi desenvolvida a partir das linguagens DAML e OIL e criada em cima da arquitetura XML e RDF. Foi proposta como padrão pelo W3C, agregando diversos pontos positivos das linguagens anteriores.

A Figura 1.2 mostra um pequeno código desenvolvido na linguagem OWL.

```

</owl:Ontology>
<owl:Class rdf:ID="CoberturaDeTomate">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:FunctionalProperty rdf:ID="temTempero"/>
      </owl:onProperty>
      <owl:someValuesFrom>
        <owl:Class rdf:ID="Suave"/>
      </owl:someValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

Figura 1.2: Código OWL
Fonte: <http://protege.stanford.edu>.

Este trecho de código em OWL mostra uma ontologia de pizzas retirado do site do Protégé.

Este quadro apresenta um comparativo entre as principais linguagens de representação de ontologias.

Linguagens	Objetivo	Vantagens
RDF	Tornar possível a especificação de	RDFs define o significado da

	regras semânticas para dados baseados em XML de maneira padronizável e interoperável. Definir um mecanismo independente de domínio. Definir um mecanismo para descrever recursos.	estrutura usando os tags estabelecidos em seu modelo. Provê tags para definir classes, subclasses e propriedades. Permite restrições de domínio e de valores (range) e propriedades.
OIL	Representação de uma semântica de informação acessível para a máquina na web.	Consegue promover pesquisas em torno do conceito semântico contido das informações das paginas web e não apenas através das palavras-chave.
DAML-OIL	Foi construída para ser uma linguagem de marcação semântica para recursos web (facilitar a representação dos conceitos de web semântica, descrevendo a estrutura de um domínio, ser compatível com padrões web existentes, fácil compreensão e uso, ter uma semântica formal associada, possuir um relativo poder de expressão)	Restrições de cardinalidade, tipos de dados, classes definidas e numerações, equivalência, extensibilidade, semântica formal, herança, inferência.
OWL	É um documento XML, RDF, RDFS Originado de vários projetos: DARPA (DAML) + Project (OIL)	Subdividido em 3 sub-linguagens: lite, DL e FULL
CYC	Proposta de uma ontologia geral com possibilidade de tratamento de contextos específicos (microteorias). Manipulação de grandes quantidades de conceitos.	Representação e tratamento de grandes ontologias suporte para inferência.

Quadro 1.2: Linguagem de representação de ontologias

3 GENE ONTOLOGY

Neste capítulo é explicado o funcionamento do consórcio Gene Ontology, assim como sua organização, estrutura e algumas ferramentas.

3.1 O que é gene ontology

É um conjunto de vocabulário controlado que definem conceitos e classificam a relação entre eles. Gene Ontology (GO) fornece vocabulário controlado para descrever o conhecimento biológico de genes e produtos genéticos.

O projeto da GO é um consórcio entre vários centros de pesquisa na área de genoma. Essa ontologia refere-se a genes e o objetivo desse projeto é fornecer um conjunto de vocabulários estruturados para um domínio biológico específico que pode ser usado para descrever produtos de genes em um organismo. (www.geneontology.org GO, 2008).

3.2 Organização da gene ontology

Os três princípios de organização da GO são componentes celulares, processos biológicos e função molecular. Um produto gene pode ser associado e localizado em um ou mais componentes celulares e é ativo em um ou mais processos biológicos, durante o qual desempenha uma ou mais função molecular (www.geneontology.org GO, 2008).

3.2.1 Componentes celulares

Um componente celular é um elemento de uma célula, mas com a ressalva de que é parte de alguns objetos maiores, esta pode ser uma estrutura anatômica (por exemplo, retículo endoplasmático ou núcleo áspero) de um gene ou grupo de produtos (por exemplo) ribossomo, (proteassoma ou uma proteína dimer).

3.2.2 Processos biológicos

Um processo biológico é uma série de eventos realizados por um ou mais conjuntos ordenados de funções moleculares (apresentados a seguir). Exemplos de processo biológicos, no sentido amplo, são *cellular physiological process* ou *signal transduction*. Exemplos mais específicos são *pyrimidine metabolic process* ou *alpha-glucoside transport*. Pode ser difícil distinguir entre um processo biológico e uma função molecular, mas a regra geral é que um processo deve ter mais do que uma etapa distinta (www.geneontology.org GO 2008).

A Figura 2.2 mostra o resultado da busca de um termo na ferramenta Amigo no site da *Gene Ontology*. Para maiores esclarecimentos sobre a ferramenta Amigo ver a seção 3.4.1.



Figura 2.2: Processo biológico

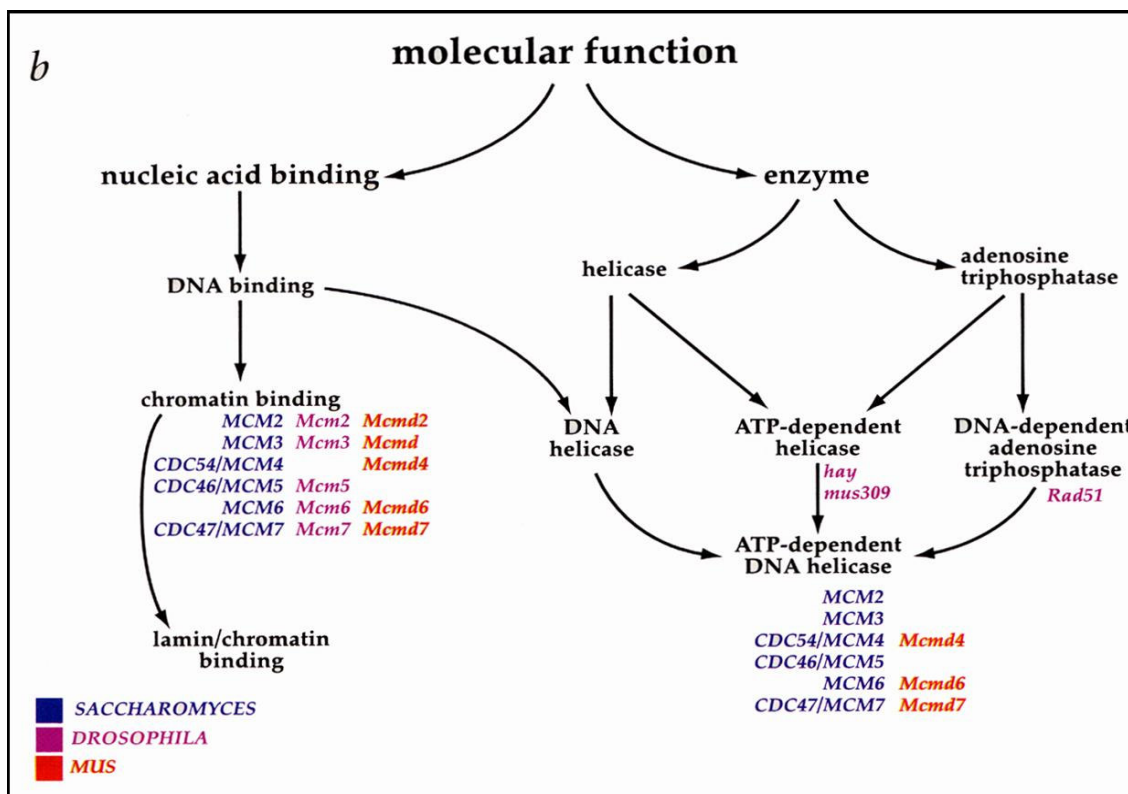
Fonte: http://gchelpdesk.ualberta.ca/news/15apr04/AmiGO/amigo_images/tree.png

3.2.3 Função molecular

Uma função molecular descreve as atividades que ocorrem em nível molecular. Exemplos de termos funcionais gerais são *catalytic activity*, *transporter activity*, e *binding*. Exemplos de termos mais restritos *adenylate cyclase activity* ou *Toll receptor binding*. O

nome do produto de um gene pode se confundido com a sua função molecular, e por esse motivo que no GO muitas funções moleculares são adicionados com a palavra “*activity*”.

A Figura 2.3 está ilustrando a estrutura da função molecular.



Figuar 2.3: Função molecular
Fonte: <http://proteinfuction.net/index.php/GO>

Este é um exemplo de uma classificação genética de um termo GO.

Nome do Gene: **BRCA1**

Função Molecular: **protein binding**

Processo Biológico: **DNA Replication and Chromosome Cycle**

Componente Celular: **nucleus**

3.3 Estrutura da gene ontology

A Figura 2.4 quem vem a seguir mostra como é a estrutura do consórcio GO, e como é definida a relação entre os conceitos. E como se relacionam os termos: “*cell*” “*nucleus*” “*membrane*”.

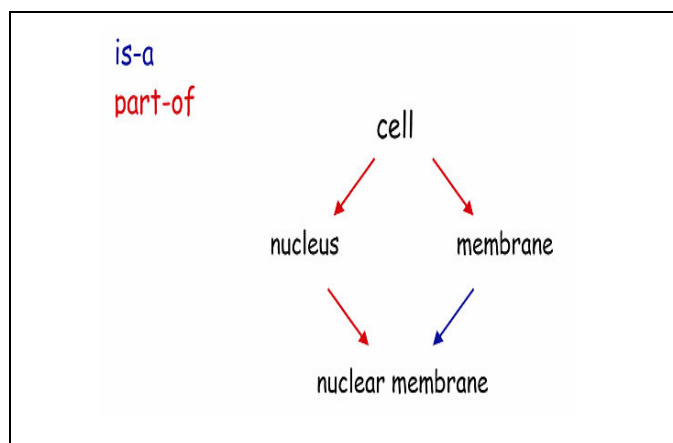


Figura 2.4: Estrutura GO
Fonte: www.geneontology.org GO 2008

3.3.1 Árvore de representação

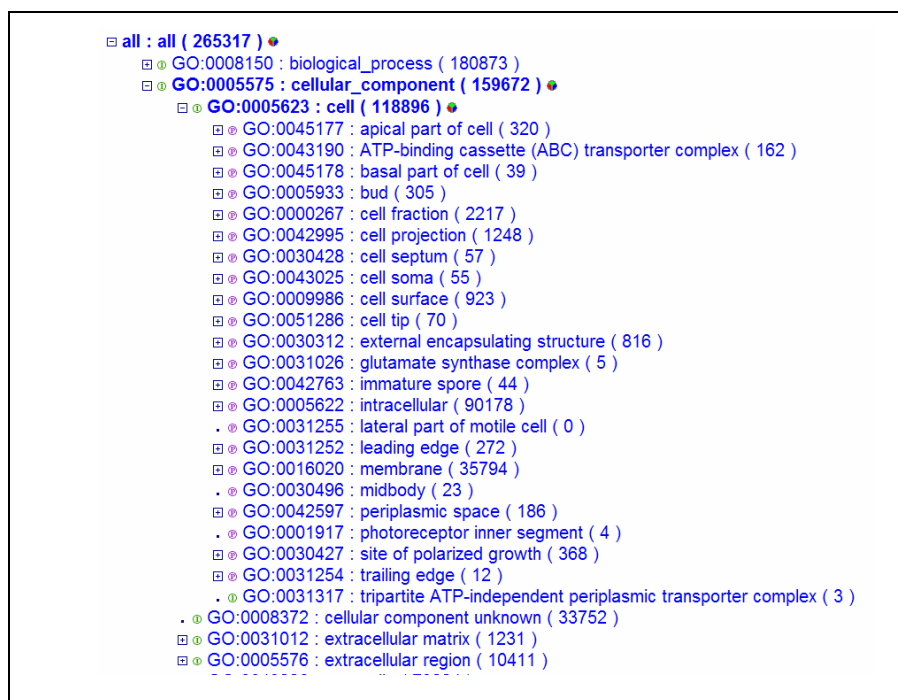


Figura 2.5: Representação da GO
Fonte: www.geneontology.org GO 2008

3.3.2 Anotação GO

Anotação é o processo de atribuição GO termos de produtos gene. A anotação de dados no banco de dados GO é feito por membros do Consórcio GO, e do Consórcio é

incentivar ativamente novos grupos a contribuem para iniciar a declaração. Esta anotação pode ser manual ou eletrônica.

Exemplo da anotação de um termo:

Symbol: SPAC22G7.07c

Name(s): mRNA

Type: gene

Species: *Schizosaccharomyces pombe* (fission yeast)

Synonyms: SPAC22G7.07c

Database: GeneDB S. pombe, GeneDB_Spombe:SPAC22G7.07c

3.4 Ferramentas GO

Atualmente no mercado existem muitas ferramentas para criação e edição de ontologias, mas para conseguir estudar bem o funcionamento delas foi selecionada duas que fazem parte do consórcio Gene Ontology e que estão relacionados com este estudo. As ferramentas escolhidas para fazer o estudo são: AmiGO e OboEdit.

3.4.1 AmiGO

A ferramenta AmiGO está disponível no site da Gene Ontology para que os usuários possam pesquisar produtos genéticos e visualizar todos os termos com os quais estão relacionados, ele também permite que você procure por um termo “RNA” e vê todos os produtos gene anotados a ele, ou procurar por um outro produto gene e vê todas suas associações. Você também pode utilizar o navegador de ontologias para ver os relacionamentos entre os termos assim como o número de produtos gene (www.geneontology.org GO, 2008).

Esta Figura 3.1 mostra a tela principal da ferramenta AmiGO.

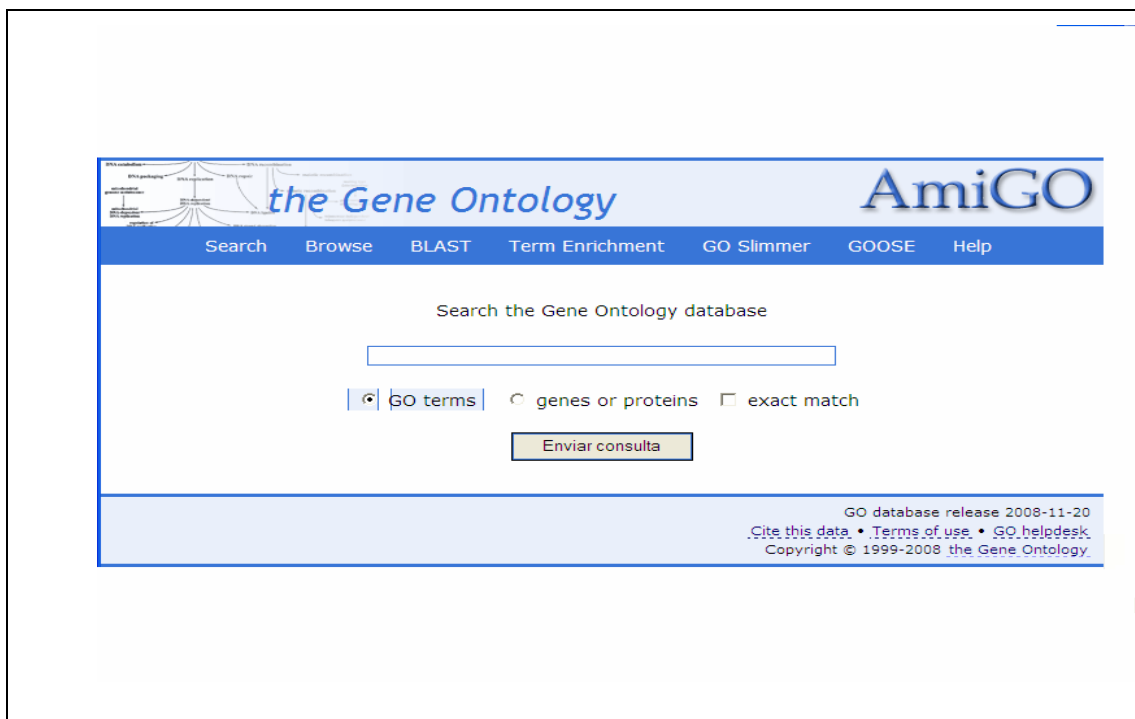


Figura 3.1: Ferramenta AmiGO
Fonte: www.geneontology.org

Com a ferramenta AmiGO é possível inserir um termo GO e ela retorna todos os produtos de gene anotados para este, ou procura por produto de gene e mostra suas associações. Na Figura 3.2 foi colocado para busca o termo “RNA” e o resultado foi de 932 associações, pode-se navegar através dos resultados, é só selecionar o termo desejado. Neste caso não foi colocado nenhum tipo de filtro, mas ele permite que você insira filtros para diminuir o número de resultados, como exemplo pode-se colocar um filtro para que o resultado mostrado seja só de termos relacionados com o componente celular ou um outro filtro em que mostre somente as associações onde aparece no nome o termo “RNA” e não nas definições. Se alguém tiver interesse em explorar essa ferramenta ela está disponível no site da GO em: <http://www.godatabase.org/cgi-bin/go.cgi>.

rel ↓	Accession , Term	Ontology
<input type="checkbox"/>	GO:0016246 : RNA interference [show def] Query matches synonym " RNA " [exact synonym]	92 gene products biological process
<input type="checkbox"/>	GO:0004526 : ribonuclease P activity [show def] Query matches synonyms " RNA se P" [exact synonym], and 1 more	64 gene products molecular function
<input type="checkbox"/>	GO:0051214 : RNA virus induced gene silencing [show def] Query matches synonyms " RNA , VIGS" [exact synonym], and 1 more	0 gene products biological process
<input type="checkbox"/>	GO:0000171 : ribonuclease MRP activity [show def] Query matches synonym " RNA se MRP" [exact synonym]	26 gene products molecular function
<input type="checkbox"/>	GO:0042245 : RNA repair [show def]	0 gene products biological process

Figura 3.2: Resultado da busca feita na ferramenta AmiGO

Após obter o resultado da pesquisa pode-se selecionar um ou mais itens para serem exibidos no formato de árvore, como mostra a Figura 3.3.

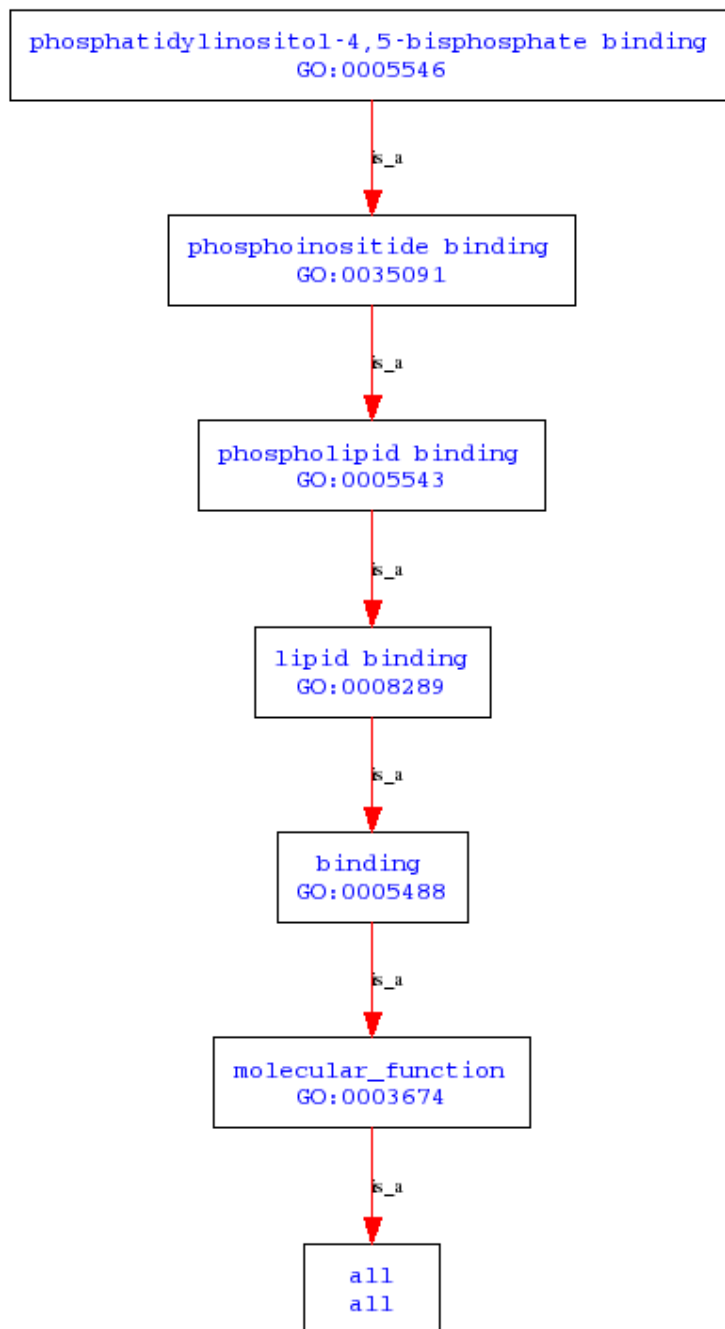


Figura 3.3: Resultado da seleção dos itens

3.4.2 OboEdit

É uma ferramenta gráfica que foi desenvolvida e é mantida pelo consórcio da Gene Ontology. Este editor foi implementado em Java, e serve para editar e visualizar ontologias. O OboEdit é particularmente muito valioso para edição e visualização de ontologias na área da biomedicina. Fornece uma interface amigável para os biólogos, e é excelente para a rápida

geração de grandes ontologias, incidindo sobre as relações entre classes relativamente simples. OboEdit por tratar-se de uma ferramenta *open source* pode ser baixada gratuitamente no site da Go, em: www.geneontology.org.

A Figura 3.4 ilustra a tela principal do OboEdit.

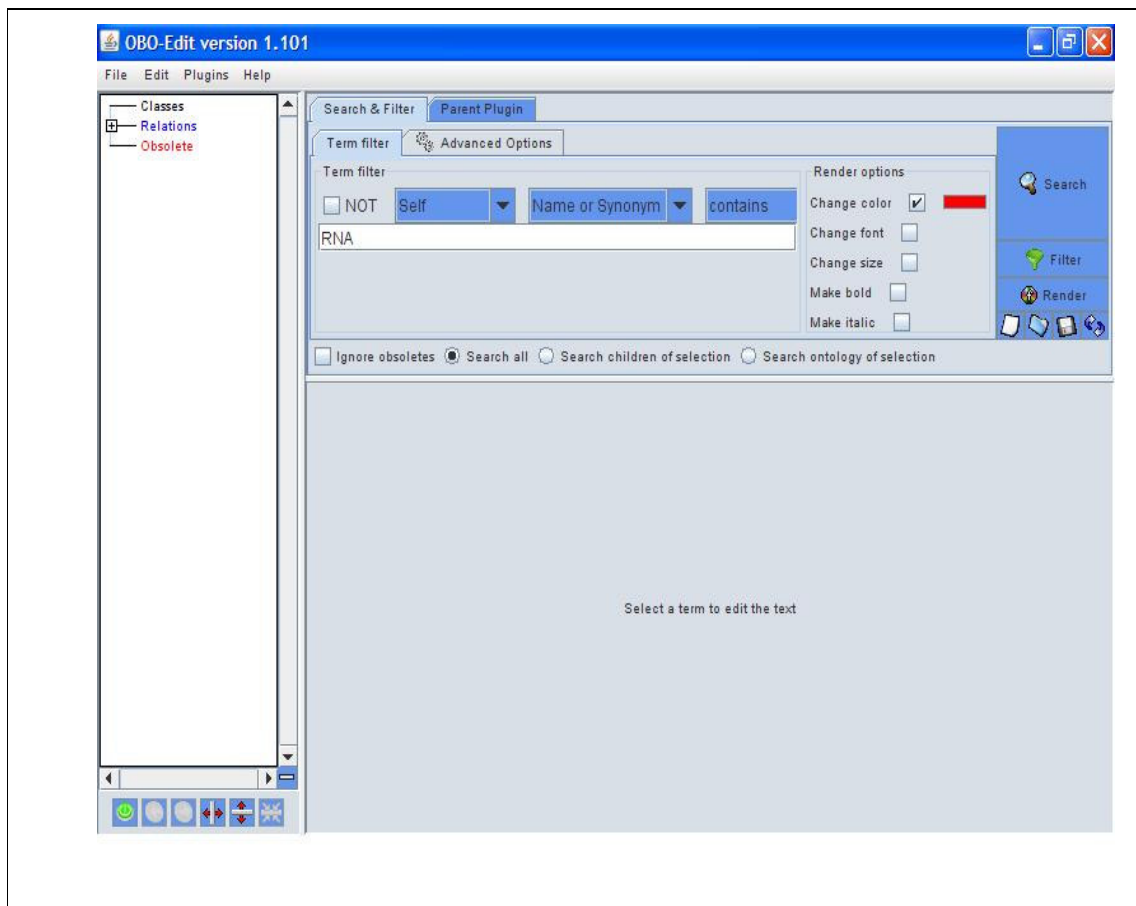


Figura 3.4: Tela principal do OboEdit

3.4.3 Anotação no OboEdit

Este exemplo mostra como os termos são descritos na ferramenta OboEdit.

```
[Term]
id: GO:0000003
name: reproduction
namespace: biological_process
alt_id: GO:0019952
alt_id: GO:0050876
def: "The production by an organism of new individuals that contain some
portion of their genetic material inherited from that organism."
[GOC:go_curators, GOC:isa_complete, ISBN:0198506732 "Oxford Dictionary of
Biochemistry and Molecular Biology"]
subset: goslim_generic
subset: goslim_pir
```

subset: goslim_plant
subset: gosubset_prok
synonym: "reproductive physiological process" EXACT []
xref: Wikipedia:Reproduction
is_a: GO:0008150 ! biological_process

4 GO SLIM

Neste capítulo é explicado como funciona o Go Slim e porque é mais fácil utilizá-lo do que usar toda a GO. Também é demonstrada a produção de um Slim utilizando a ferramenta OboEdit.

4.1 O que é Go Slim

É um subconjunto da GO, particularmente são úteis para dar um resumo dos resultados da anotação GO de um genoma. GO Slims são criados pelos usuários de acordo com suas necessidades, e podem ser específicas ou para uma área particular da ontologia.

4.2 Por que usar Go Slim

Go Slim é usado simplesmente pelo motivo de que usar toda a GO não é satisfatório, pelo grande volume de termos a ela associados. Os Slims são gerados a partir de uma parte da GO ou também de outros Slims. Podem-se pegar as anotações de uma GO plantas e a partir dessa gerar um novo Slim específico de uma determinada espécie de planta.

4.3 Produzindo um Slim utilizando o OboEdit

Em razão da documentação sobre como construir Slims ser escassa e simplificada, é desenhado a seguir os passos para executar esse procedimento.

Pode ser utilizado um dos exemplos de Slims baixados do site da GO. Após abrir o OboEdit selecione no menu file a opção: *Load Terms*, em seguida aparecerá uma janela onde deverá ser selecionado o tipo de adaptador, selecione *OBO Flat File Adapter*, após utilize a opção *browse* para selecionar o slim que efetuará a busca, clique em OK e o arquivo será carregado. Na opção *Term filter* insira um termo para efetuar a busca em seguida clique em *Search*, caso seja encontrado algum termo relacionado com este aparecerá na árvore à

esquerda, caso queira realçar os termos, pode marcar na opção *Render options* o item *Change color* com a cor desejada para os termos. Em seguida selecione os termos que quer nos resultados encontrados e clique em *filter*. Este filtro pode ser salvo para que possa utilizá-lo em outros slim. O filtro pode ser salvo no disquete que aparece logo abaixo da opção *Render* em *Save a filter to disk* e nesse mesmo item é possível carregar os filtros na opção ao lado *Load a filter from disk* ou criar novo filtro em *Create new filter*.

Após aplicar o filtro, pode-se salvar esse slim junto com o filtro criado na opção *file* → *Save as*. Para testar o slim criado abra ele no OboEdit.

4.3.1 Produzindo um Slim Passo a passo

As Figuras mostram como selecionar um Slim e a partir deste inserir um termo para pesquisar e com o resultado da busca selecionar alguns e salvar o novo Slim.

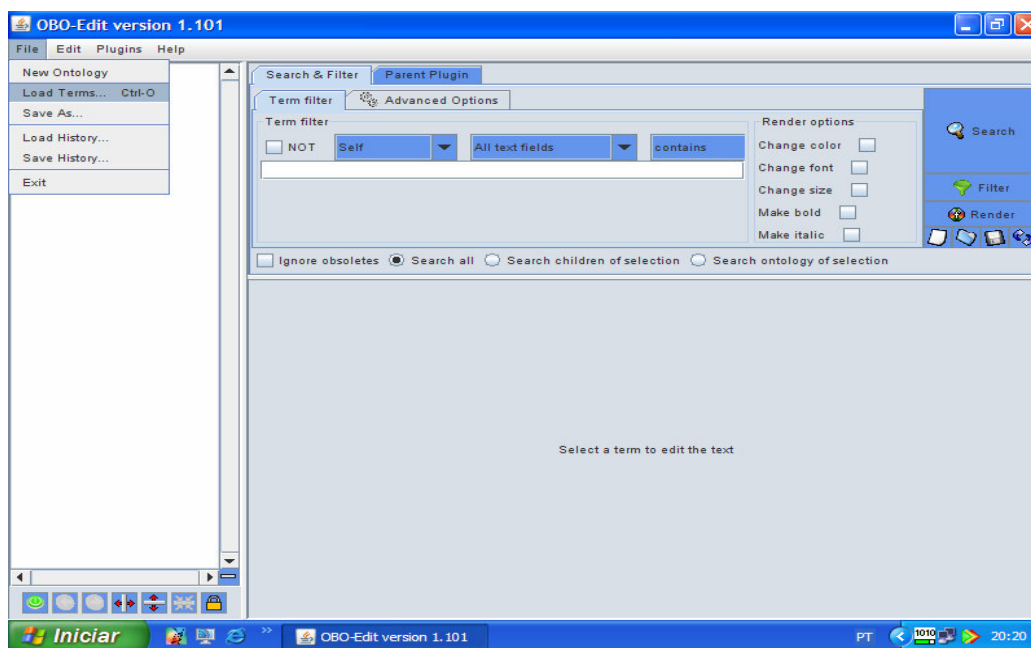


Figura 4.1 – Busca arquivo para análise.

Esta Figura ilustra como faz para carregar ontologias na ferramenta OboEdit. Através do menu *File* opção *Load Terms*.

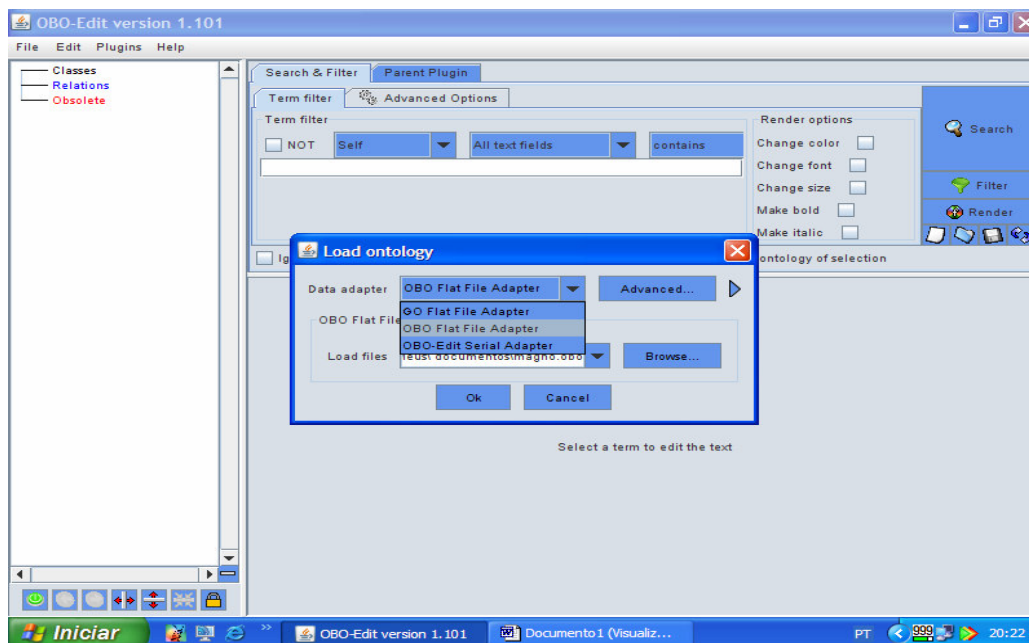


Figura 4.2 – Selecionar tipo do adaptador.

A Figura 4.2 mostra quais os tipos de adaptadores disponíveis na ferramenta. Neste exemplo foi selecionado o adaptador *OBO Flat File Adapter*.

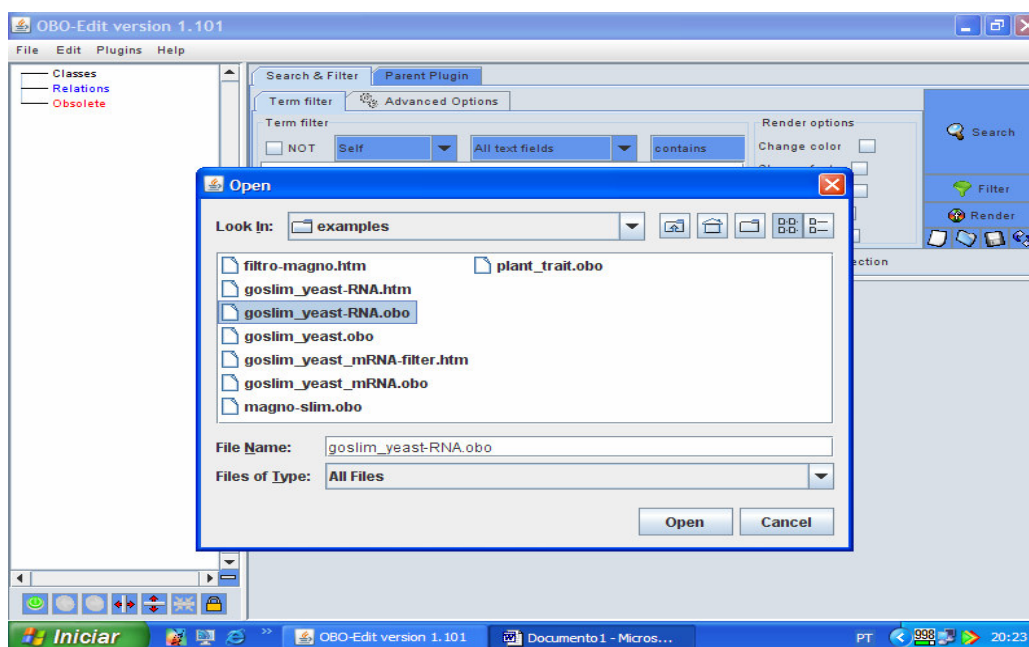


Figura 4.3 – Seleção do arquivo na pasta exemplos do OboEdit.

Na Figura 4.3 esta sendo mostrado qual arquivo foi selecionado para fazer a demonstração da pesquisa com o termo “RNA”.

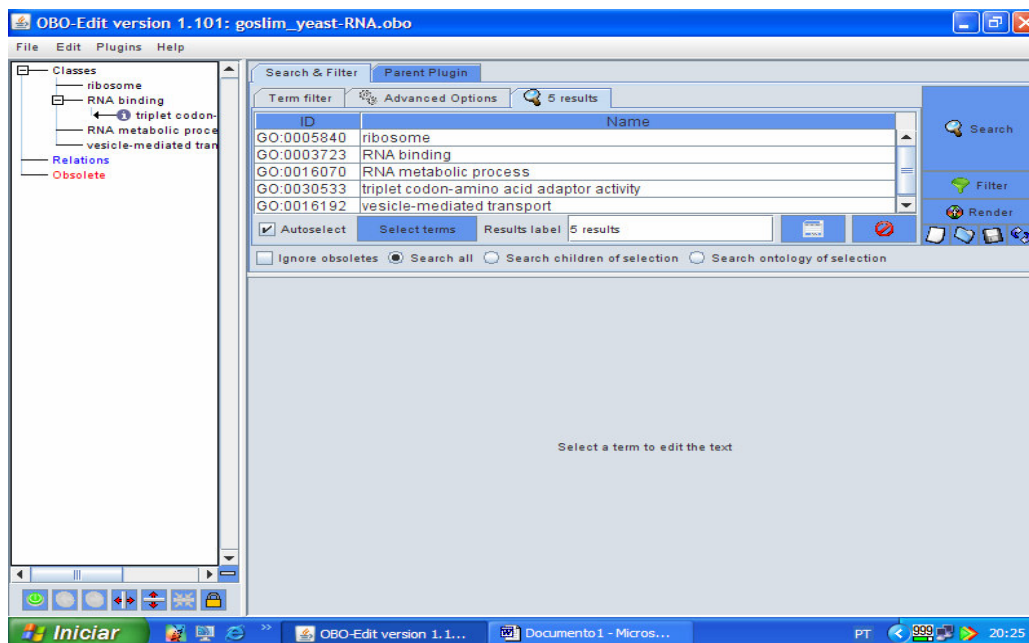


Figura 4.4 – Inserir o termo para pesquisa

A Figura 4.4 mostra o retorno após aplicar a busca no termo “RNA”, que obteve 5 associações.

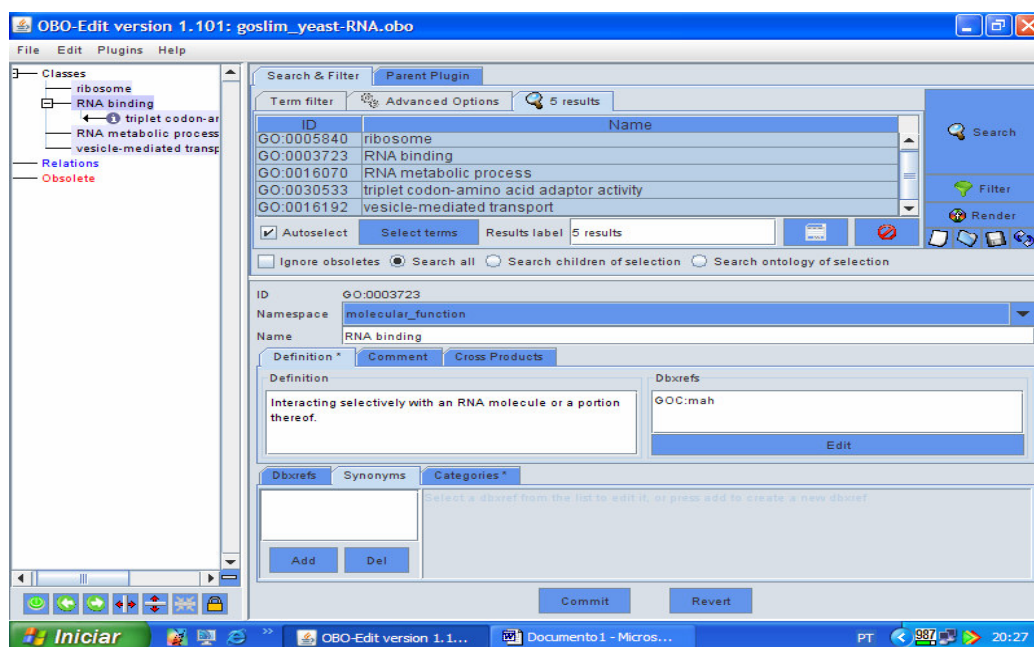


Figura 4.5 – Selecionar os termos

Nesta Figura esta em destaque os termos selecionados para o filtro, que ao serem selecionados ficam com fundo azul.

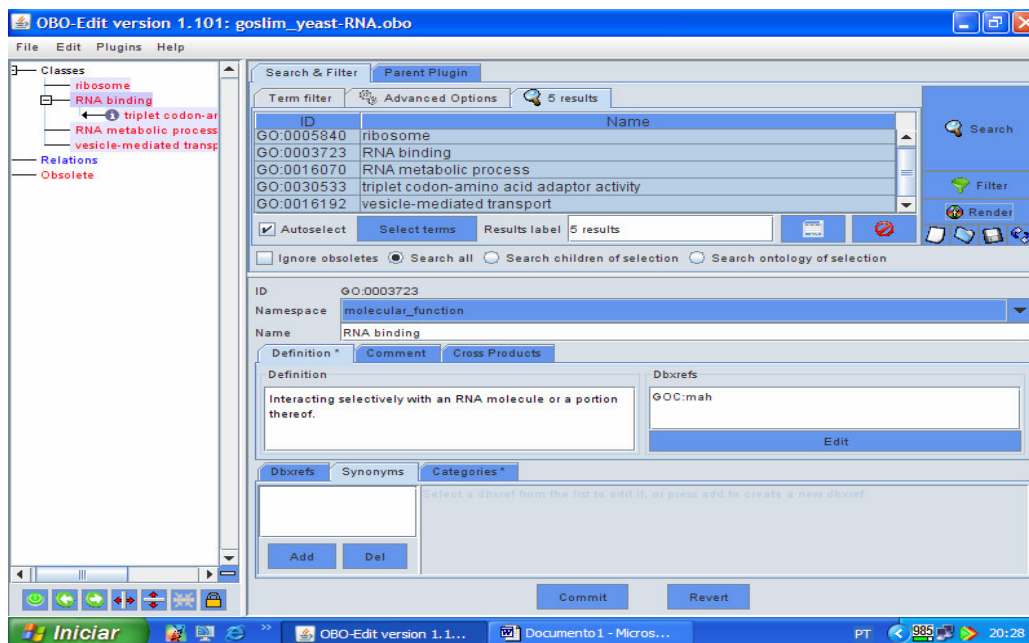


Figura 4.6 – Mudar cor dos itens

A Figura 4.6 mostra como fazer para mudar a cor dos itens selecionados para compor o filtro. Para ser executado esse processo, você deve clicar na aba *Term filter* e marcar a opção *Change color* e ao lado escolher a cor conforme desejado, neste exemplo foi selecionada a cor vermelha.

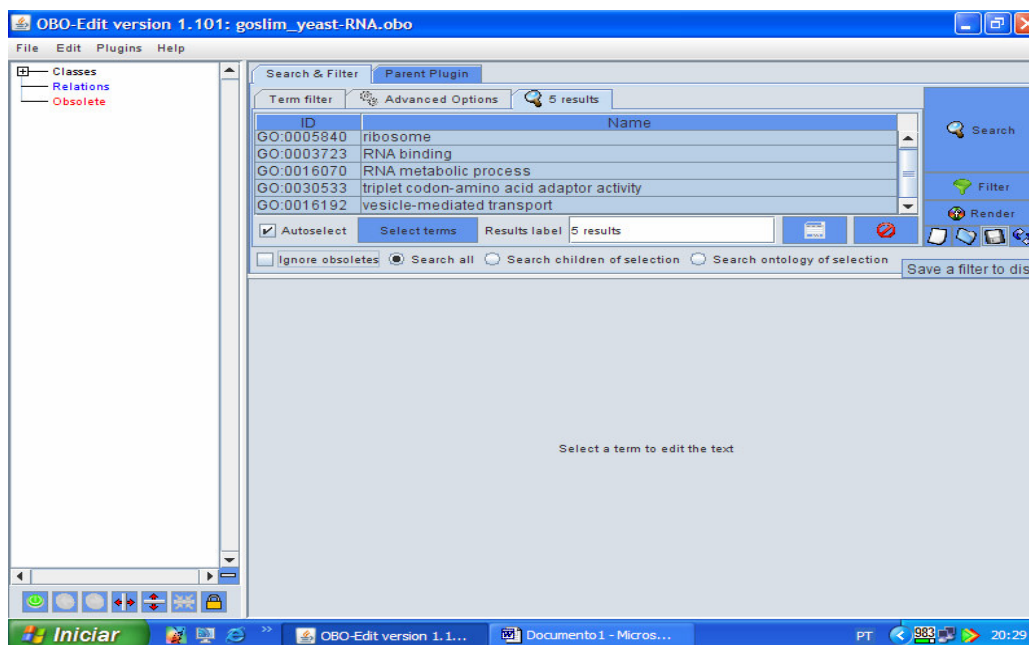


Figura 4.7 – Como criar e salvar um filtro

No OboEdit pode-se criar e salvar um filtro seguindo esta dica: primeiro clicar em *Filter* para criá-lo conforme Figura 4.7 e depois selecionar a opção para salvar o filtro. Caso queira salvar o Slim, clique no menu *Filte* e depois *Save As*, conforme Figura 4.8.

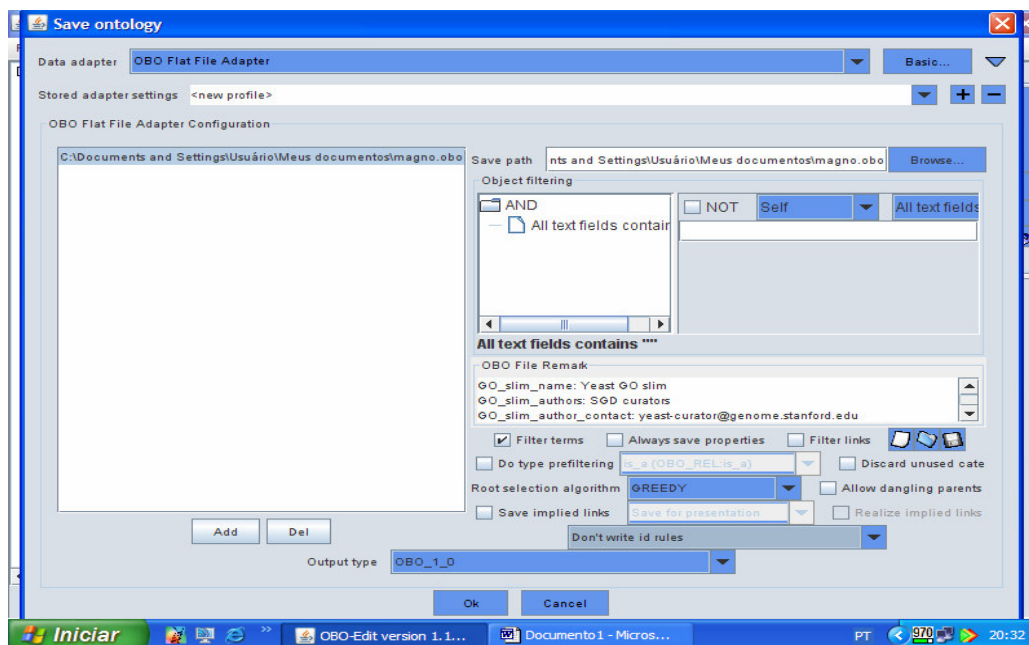


Figura 4.8 – Salvar o Slim com o filtro

A Figura 4.8 ilustra como salvar junto com o Slim o filtro. Em *Save path* é inserido o caminho e o nome do Slim. Para salvar o filtro junto você deve carregar o filtro na opção que aparece ao lado da caixa de seleção *Filter links*, na opção salvar pode ser criado novo filtro.

4.3.2 Editando um Slim

Após a criação do Slim, ele pode ser alterado, incluindo novos termos ou alterando os que já existem. Esta alteração pode ser nas definições do termo como por exemplo acrescentar sinônimos e apontar a qual base de dados que ele pertence.

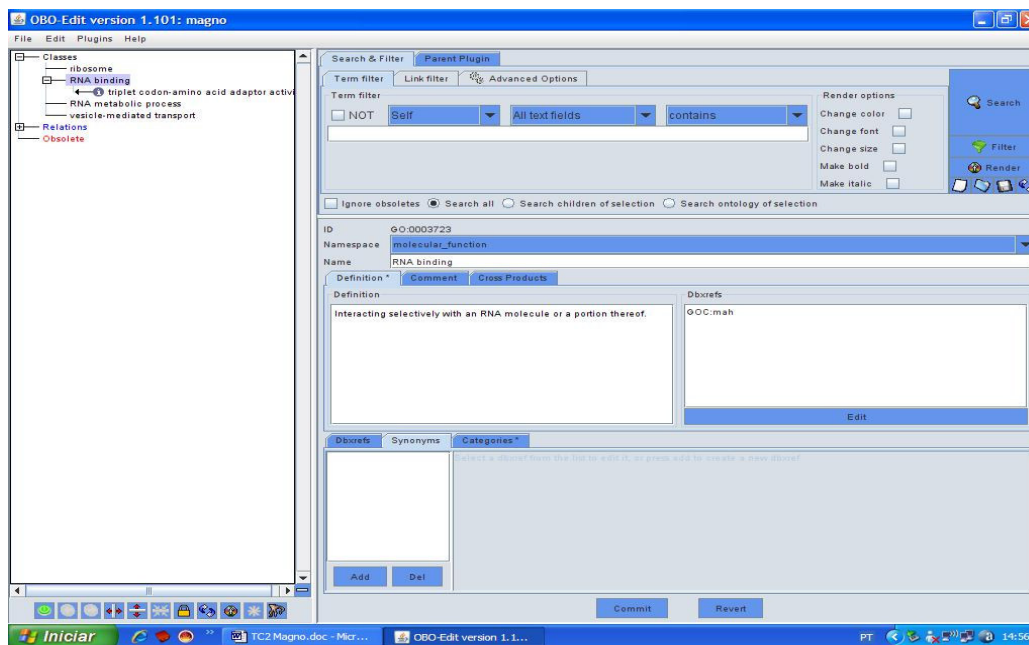


Figura 4.9 – Mostra Slim criado para teste

A Figura 4.9 mostra o Slim (magno.obo), que foi criado no OboEdit para teste, no qual foram selecionados quatro termos da pesquisa realizada com a palavra chave “RNA”.

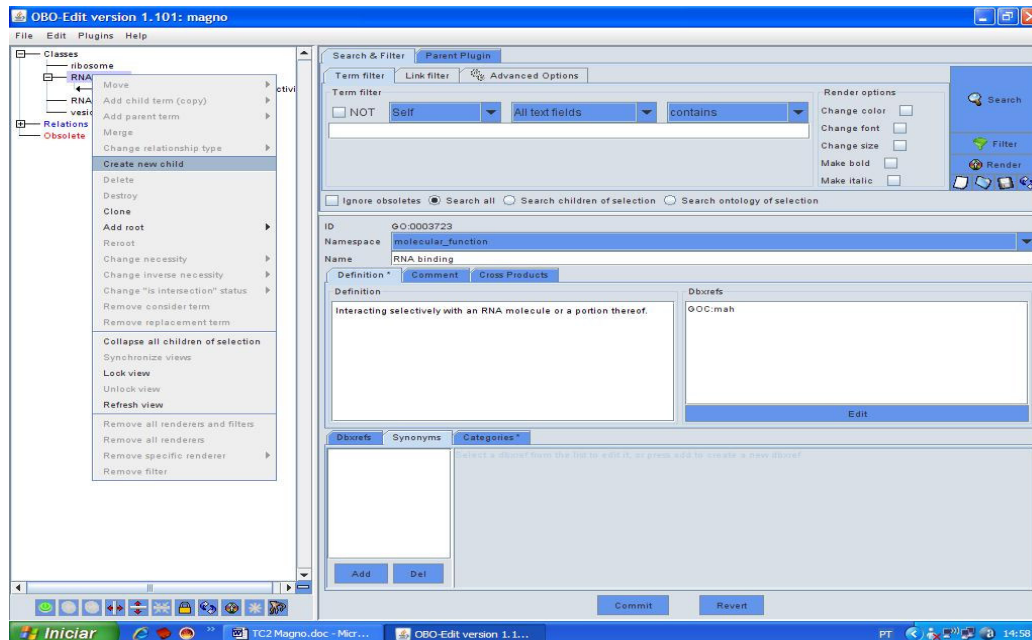


Figura 4.10 – Criando um novo termo

Na Figura 4.10 podemos ver a ilustração de como criar um termo filho para o termo maior denominado pai. Neste exemplo está sendo criado um filho para o termo pai “RNA binding”.

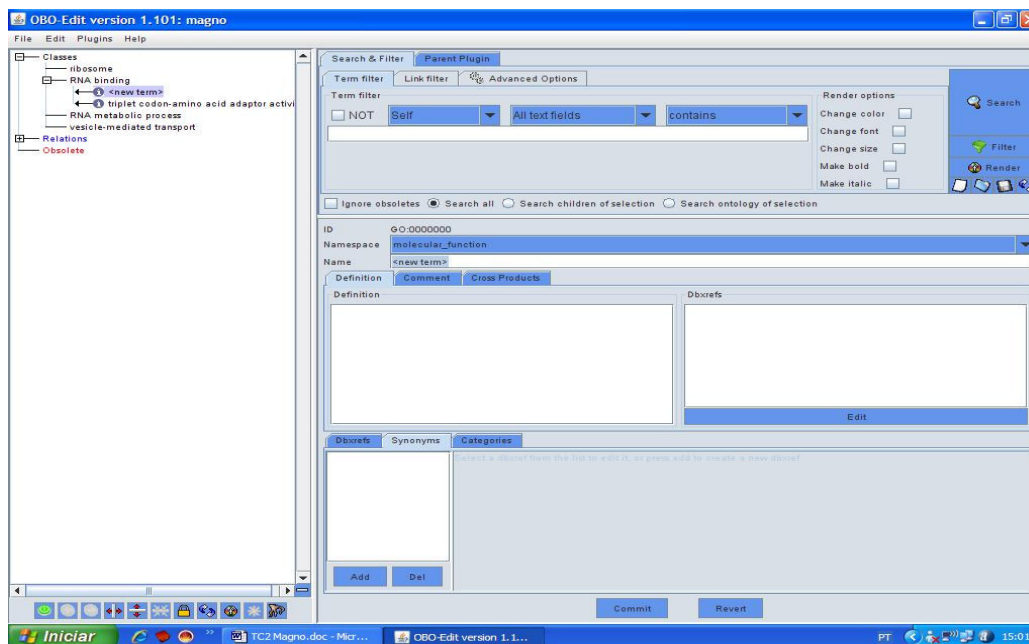


Figura 4.11 – Atribuindo nome ao novo termo

Após ter criado o novo termo e dado seu nome como mostra a Figura 4.11, também podemos definir se ele é um processo biológico, função molecular ou um componente celular, assim como descrever a sua definição e apontar de onde foi tirada essa informação e colocar seus sinônimos.

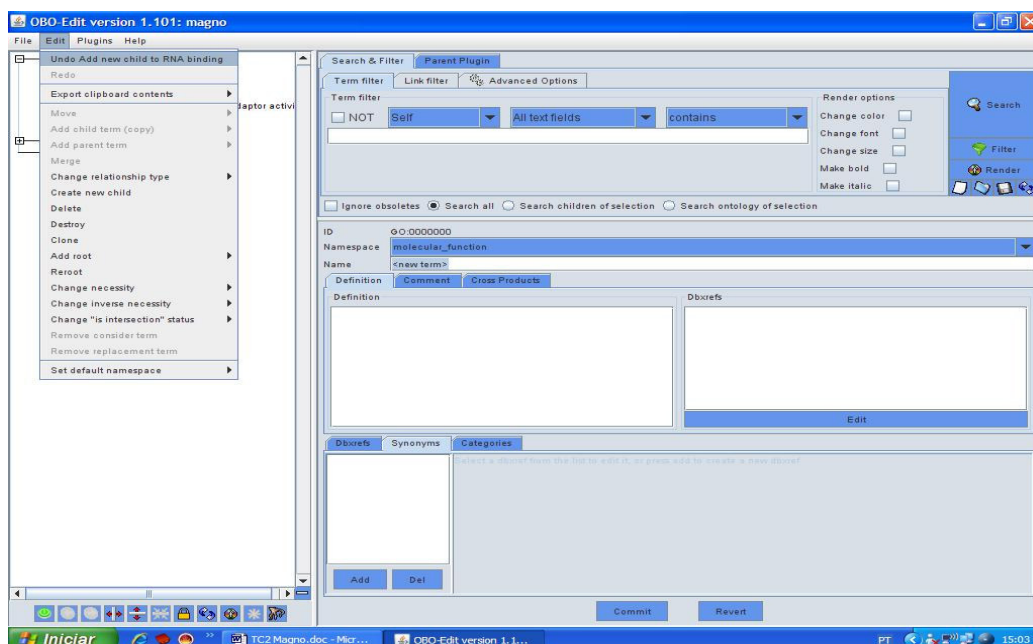


Figura 4.12 – Desfazendo as alterações

Todas as alterações efetuadas na ferramenta podem ser desfeitas através da opção do menu *Edit*, item *Undo*, como está ilustrada nesta Figura 4.12.

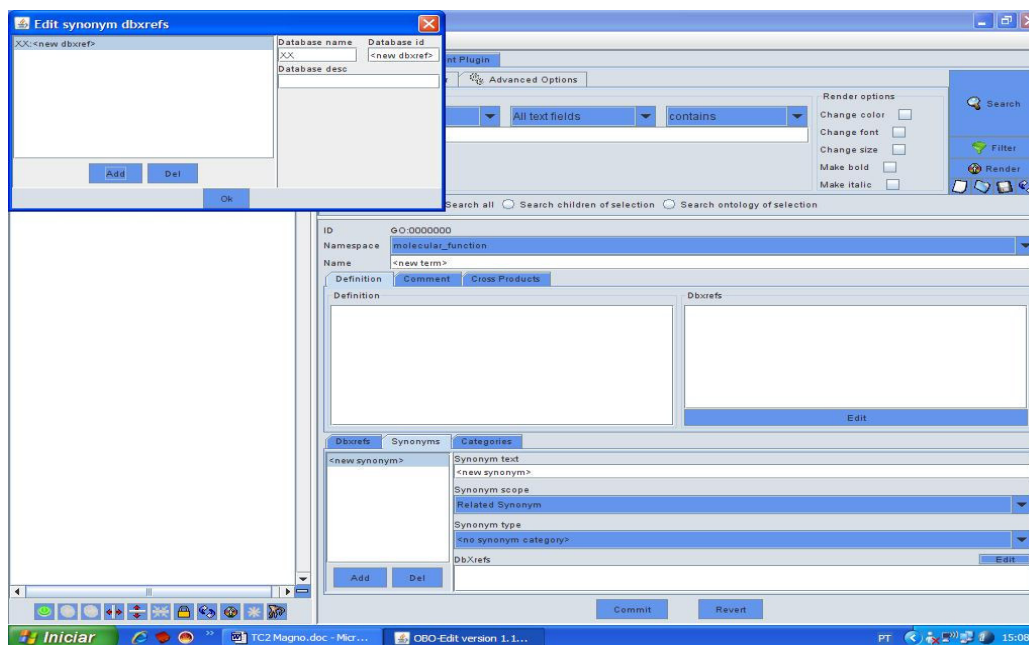


Figura 4.13 – Inserindo Base de Dados

Podemos verificar na Figura 4.13 como é feita a inserção da base de dados onde ficam disponível os artigos contendo informações mais detalhadas sobre o termo que está sendo inserido nesta ontologia. O OboEdit permite que sejam inseridas mais de uma base de dados contendo informações deste termo.

4.3.3 Considerações sobre o OboEdit

Nos testes realizados com o OboEdit foi constatado que é uma ferramenta que para gerar ontologias deve-se conhecer muito ela. Claro que através do *help* é possível gerar ontologias, seguindo as instruções passo a passo, mas que para isso você tem que saber pelo menos um pouco de inglês para conseguir utilizar a ferramenta. Após criar as ontologias você pode testá-las abrindo no OboEdit. Depois de entender o funcionamento dela, torna-se fácil e muito útil para a criação e manutenção de ontologias. Uma vantagem importante que deve ser considerada, é que o OboEdit está disponível tanto em plataforma Windows como também em Linux.

5 SLIM PARA PESQUISA EM HIV

5.1 Introdução

Após ter realizado pesquisas sobre a geração de Slims voltado para a biologia, que foi o tema descrito no TC1, o estudo foi apresentado ao grupo de pesquisa especialista da Feevale, das áreas de biologia e genética, para verificar os problemas enfrentados na pesquisa, e como e quanto este estudo poderia de alguma forma auxiliar o desenvolvimento do mesmo.

O encontro primeiramente foi com a coordenadora do grupo de pesquisa a Prof. Dra. Sabrina Esteves de Matos Almeida que possui graduação em Ciências Biológicas pela Universidade Federal do Rio Grande do Sul (1996), mestrado em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (1999) e doutorado em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (2003). Atualmente é professora adjunta da Feevale e funcionária da Fundação Estadual de Produção e Pesquisa em Saúde. Tem experiência na área de Genética, com ênfase em Biologia Molecular e de Microorganismos, atuando principalmente nos seguintes temas: Bioinformática aplicada a microorganismos patogênicos, Metodologias aplicadas ao diagnóstico molecular e Genética humana e animal, que nos explicou como funciona o grupo e demonstrou interesse em colaborar com nosso projeto, pois o projeto seria muito útil para eles, porque segundo ela gerando esses Slims de domínios específicos serviria tanto para uso em sala e também para o grupo de pesquisa.

A pesquisadora utiliza com seus alunos e pesquisadores o site da NCBI (<http://www.ncbi.nlm.nih.gov>), para efetuar suas pesquisas e segundo ela, principalmente, os iniciantes tem alguma dificuldade em utilizar esse mecanismo de pesquisa, por ainda não terem um conhecimento sobre os temas e o funcionamento do site de pesquisa. Por estes alunos não terem o conhecimento sobre o que estão pesquisando o retorno é muito grande de artigos que nem sempre tem relação com o tema pesquisado.

5.2 Proposta

Dos conhecimentos já adquiridos pelos alunos e pesquisadores, montar Slims contento todas as informações necessárias referentes ao produto pesquisado e a partir destes fazer uma breve descrição e inserir links com informações de livros e artigos que foram pesquisados sobre esse tema facilitando futuras pesquisas.

Para ilustrar como gerar um Slim foi desenvolvido, em conjunto com o grupo de pesquisa em Bioanálises, uma Ontologia do domínio HIV com suas classes, subclasses e instâncias. As informações necessárias para criar esse Slim foram passadas pela aluna do curso de Biomedicina, Nicole Pezzi que também faz parte do Grupo de Pesquisa em Bioanálises da Feevale.

A especialista reuniu informações referentes aos genes e proteínas para criação da ontologia, em função destas informações foi utilizado a ferramenta OboEdit na montagem deste Slim.

5.3 Aplicação

Em sala de aula como material de apoio aos professores e também para tornar mais fácil a pesquisa dos alunos uma vez que podem buscar as informações mais específicas sem precisarem dominar o tema. Os Slims serão utilizados e mantidos por professores, alunos, pesquisadores na área da Biologia e por pessoas interessadas pelo assunto.

5.4 Material disponibilizado pelo especialista

Com o material disponibilizado foi utilizado o OboEdit para verificar se os termos já existiam, com os resultados obtidos montar um slim, caso algum destes itens não fossem encontrados pela ferramenta os mesmos seriam incluídos no slim criado.

A Figura 5.1 mostra os termos que a especialista disponibilizou para a pesquisa.

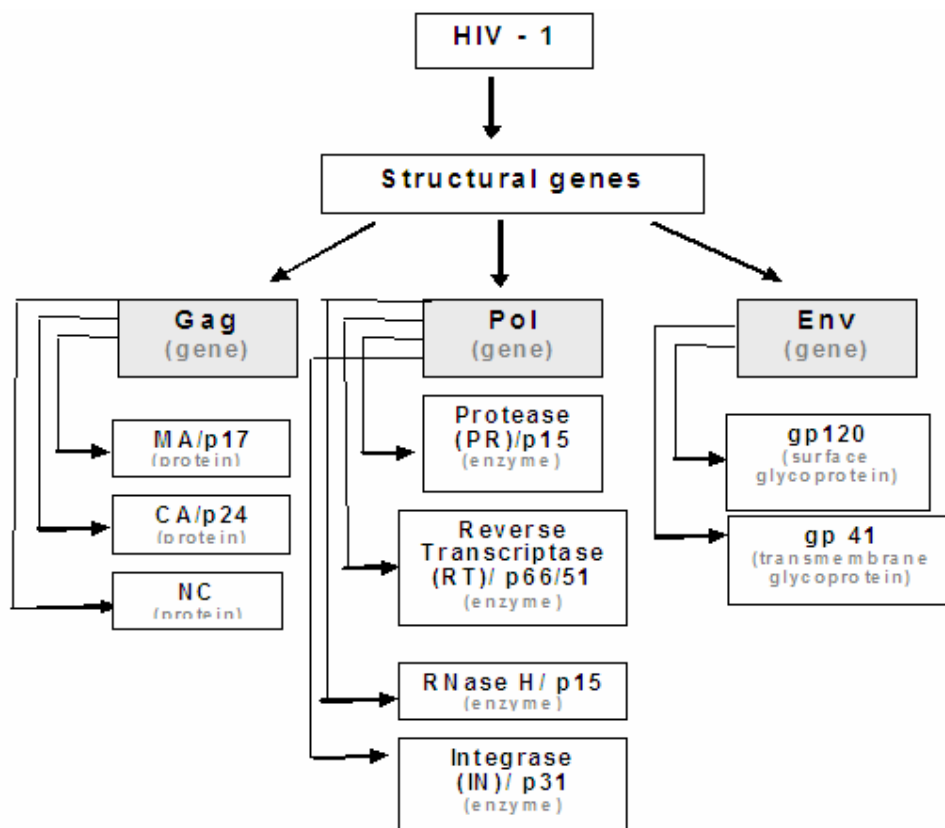


Figura 5.1 – Material disponibilizado pela especialista

5.5 Genes e produtos gene

Definição segundo o site: <http://decs.bvsalud.org/cgi-bin>

Gag: Proteínas são codificadas pelo gene gag retroviral. Os produtos, geralmente são sintetizados como precursores de proteínas ou POLIPROTEÍNAS, que são divididas por proteases virais para liberar os produtos finais. Muitos dos produtos finais estão associados com a proteína nuclear do vírus. O gag é curto para o antígeno específico do grupo.

Pol: DNA polimerase dependente de DNA, caracterizada em procariotos, e que pode estar presente em organismos superiores. Tem tanto atividade de exonuclease 3'-5' quanto 5'-3', mas não pode usar o DNA de fita dupla nativo como molde-iniciador. Não é inibida por reagentes sulfidrílicos e é ativa tanto na síntese quanto no reparo do DNA.

Env: Proteínas retrovirais, freqüentemente glicosiladas, codificadas pelo gene envelope (env). Geralmente, são sintetizados como precursores de proteínas

(POLIPROTEÍNAS) e mais tarde divididos em glicoproteínas de envelope viral por uma protease viral.

Segundo a especialista os genes são associados à função molecular enquanto as proteínas são associadas a um processo biológico e também podem ser associados ao componente celular, enquanto as enzimas são associadas somente ao processo biológico.

5.6 Criação do Slim

Os termos sugeridos pela especialista foram pesquisados na ferramenta OboEdit em todos os campos possíveis (nomes de genes, proteínas, descrição, etc). Daqueles que apresentaram algum resultado, foram selecionados os de maior interesse e então foram adicionados a um novo slim. A busca ocorre da seguinte forma:

- 1) na ferramenta é inserido o termo como por exemplo o gene GAG,
- 2) dos resultados apresentados, são filtrados somente aqueles que tiveram relação com a função molecular (em razão de o termo GAG ser um gene).

Primeiramente, é carregado todo a GO na ferramenta OboEdit, no menu *file* na opção *Load Terms*, em seguida inserir no campo *Term filter* o termo para a busca, neste caso vamos utilizar o termo GAG, após pressionar o botão *Search* será exibido os resultados, como mostra a Figura 5.2 abaixo no campo *Term results*.

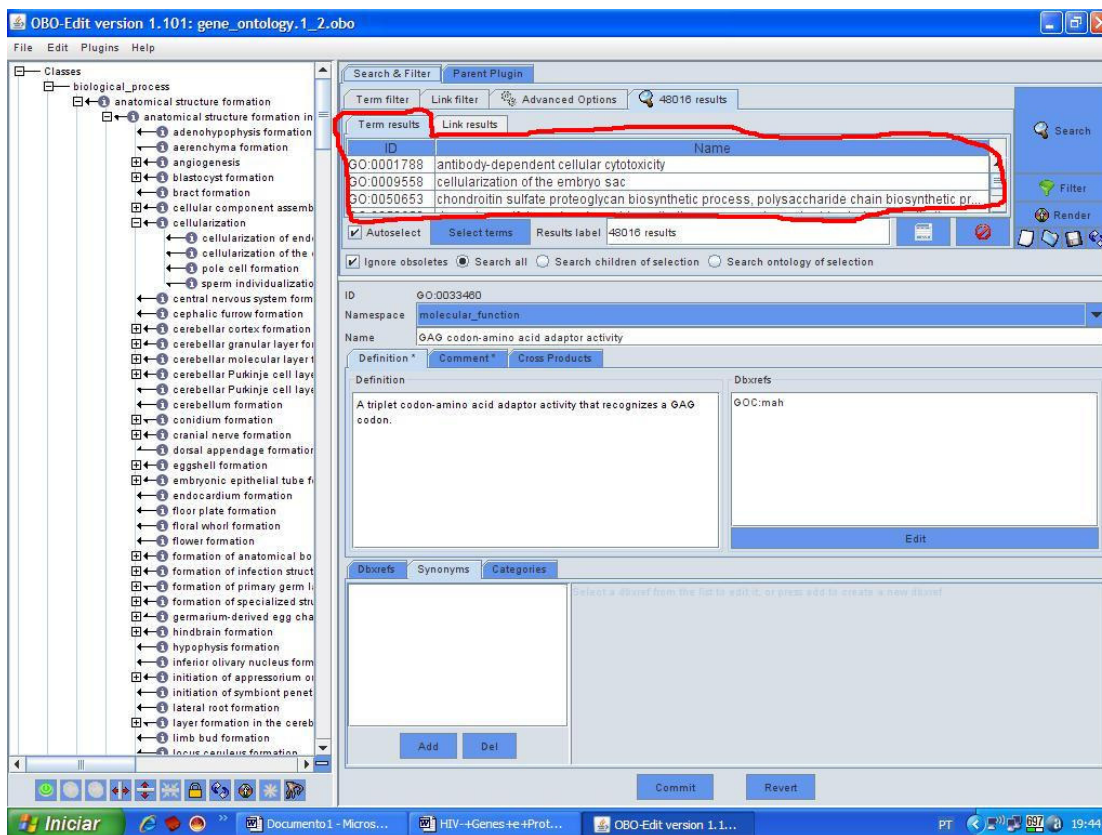


Figura 5.2 – Resultado da busca pelo termo Gag.

Após obter os resultados da pesquisa, verificar se cada um deles tem relação com o termo pesquisado e em seguida ver se é uma função biológica segundo a especialista. No caso da Gag apenas um termo foi selecionado para compor o slim, o gene Pol teve mais termos selecionados e o gene Env teve também apenas um termo selecionado como está demonstrado abaixo.

Gene: Gag

Termo selecionado:

ID: 0033460

Namespace: molecular_function

Name: GAG codon-amino acid adaptor activity

Gene: Pol

Termo selecionado:

ID: 0046428

Namespace: molecular_function

Name: 1,4-dihydroxy-2-naphthoate octaprenyltransferase activity

ID: 0043430

Namespace: molecular_function

Name: 2-decaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

ID: 0043428

Namespace: molecular_function

Name: 2-heptaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

ID: 0043434

Namespace: molecular_function

Name: 2-hexaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

ID: 0043429

Namespace: molecular_function

Name: 2-nonaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

ID: 0043433

Namespace: molecular_function

Name: 2-octaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

ID: 0008425

Namespace: molecular_function

Name: 2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

ID: 0033937

Namespace: molecular_function

Name: 3-deoxy-2-octulosonidase activity

ID: 0050534

Namespace: molecular_function

Name: 3-deoxyoctulosonase activity

Gene: Env

Termo selecionado:

ID: 0004673

Secondary IDs: 0008896

Namespace: molecular_function

Name: protein histidine kinase activity

O mesmo procedimento é adotado na pesquisa de proteínas e enzimas, assim como os genes os resultados selecionados são adicionados ao slim. Esse slim pode ser revisado e atualizado conforme necessidade.

CONSIDERAÇÕES FINAIS

No decorrer deste trabalho foi observado o quanto às ontologias são úteis para o desenvolvimento do conhecimento, seja ela aplicada na área da biologia ou em outras áreas, com a evolução da informática e da Internet ficou mais fácil desenvolver esse projeto uma vez que se consegue atingir cada vez mais grupos interessados a contribuir com a evolução das ontologias.

De certa forma a ontologia ajuda a não ter que criar sempre do zero algum projeto, uma vez que alguém já tenha criado algo que pode ser utilizado, a partir desse projeto gerar um novo, seria como se sempre tivesse que reinventar a roda. Na biologia, não é diferente. Poderá utilizar ontologias já criadas e a partir dessas, gerar novas e sempre deixar a disposição de outras pessoas que queiram utilizá-las como fonte de estudo para criar outras ontologias.

O que mais chama atenção é a dificuldade encontrada pelos biólogos e pessoas interessadas em pesquisar sobre o tema produtos genéticos que é a terminologia que é dada aos produtos. Por exemplo, a proteína “p24” numa base de dados pode ser anotada como “p24-Gag” em outra é anotada como “p24” ou simplesmente “CA”, isto ocorre com a maioria das anotações tornando difícil o entendimento. Por este motivo a necessidade de criar ontologias onde relaciona-se os termos que tem a mesma definição mas com terminologia diferente, no caso da “p24”, coloca-se como nome específico p24 os demais como sinônimos já que se trata da mesma proteína.

No estudo que foi realizado, procurou-se o auxílio de um grupo de pesquisa, onde eles disponibilizariam dados referente a genes e proteínas de uma espécie. Neste caso, foi disponibilizado a espécie do HIV, foram selecionados alguns genes e suas proteínas para submeter a uma ferramenta de geração de ontologias. Após a aplicação foram separados somente os termos que tinham relação com o que se estava pesquisando e a partir destas informações gerou-se uma ontologia específica. A idéia é sempre gerar ontologias menores

para facilitar as pesquisas, sempre é mais fácil encontrar uma informação sabendo onde que ela está. Seria como ir a um lugar que a gente não tem referência nenhuma, é possível chegar mas vai demorar bem mais do que já tivéssemos em mãos cada passo a ser dado até chegar ao destino.

O projeto *Gene Ontology* (GO) é um grande repositório de ontologias, para quem desejar aprofundar-se mais sobre o assunto ontologia na biologia. É sugerido que consulte o site da GO em: www.geneontology.org.

REFERÊNCIAS BIBLIOGRÁFICAS

BREITMAN, Karin Koogan. **Web Semântica a Internet do Futuro**. Editora LTC, Rio de Janeiro, 2005.

HINZ, Verlaní Timm. **Proposta de Criação de uma Ontologia de Ontologias**. Disponível em: <http://ppginf.ucpel.tche.br/TI-arquivos/2006/VerlaníHinz/PPGINF-UCPel-TI-2006-2-10.pdf>. Acessado em: 05/09/2008.

Gene Ontology. **The Gene Ontology**. Disponível em: <http://www.geneontology.org>. Acessado em: 18/08/2008.

SMITH, Barry. **Ontology and Information Systems**. [s.l.: s.n.], 2003. Disponível em: < <http://ontology.buffalo.edu/smith/articles/ontologies.htm> >. Acesso em: 20/10/2008.

Gene Ontology. **What is Ontology**. Disponível em: http://www.geneontology.org/teaching_resources/presentations/2007-11_lungmeeting_dhill.ppt. Acessado em: 03/11/2008

W3C. **OWL Web Ontology Language Use Cases and Requirements**. Disponível em: <http://www.w3.org/TR/webont-req/#onto-def>. Acessado em: 05/09/2008.

The Semantic Web Vision. Disponível em: <http://mitpress.mit.edu/books/chapters/0262012103chap1.pdf>. Acessado em: 04/09/2008.

LGE. **Bancos de Dados de Genomas**. Disponível em: http://www.lge.ibi.unicamp.br/pagina/estrutura/montagem_script/exibe_ferramenta_pagina.php. Acessado em: 11/11/2008.

BECKER, Júnior. **Ontologia Lingüística para apoio a Ferramentas de Recuperação de Informações e de Text Mining**. Disponível em: <http://ead.feevale.br/tc>. Acessado em: 13/11/2008.

MATOS, Ely Edison da Silva. **Um Framework baseado em Ontologias com Serviços Web para Modelagem Conceitual em Biologia Sistêmica**. Disponível em: http://www.ufjf.br/mmc/files/2008/09/dissertacao_ely_matos.pdf. Acessado em: 14/11/2008.

DeCS. Descritores em Ciências da Saúde. Disponível em: <http://decs.bvsalud.org/cgi-bin>. Acessado em : 13/06/2009.

ANEXO – SLIM HIV GERADO NO OBOEDIT

format-version: 1.2
date: 18:05:2009 15:31
saved-by: dph
auto-generated-by: OBO-Edit 2.0
subsetdef: goslim_candida "Candida GO slim"
subsetdef: goslim_generic "Generic GO slim"
subsetdef: goslim_goa "GOA and proteome slim"
subsetdef: goslim_pir "PIR GO slim"
subsetdef: goslim_plant "Plant GO slim"
subsetdef: goslim_pombe "Fission yeast GO slim"
subsetdef: goslim_yeast "Yeast GO slim"
subsetdef: gosubset_prok "Prokaryotic GO subset"
synonymtypedef: systematic_synonym "Systematic synonym" EXACT
default-namespace: gene_ontology
remark: cvs version: \$Revision: 1.615 \$

[Term]

id: GO:0008425
name: 2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity
namespace: molecular_function
def: "Catalysis of the reaction: 2-polyprenyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-methionine = 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-homocysteine." [GOC:kd, PMID:9083048]
subset: gosubset_prok
synonym: "2-polyprenyl-6-methoxy-1,4-benzoquinone methylase activity" EXACT []
synonym: "coenzyme Q biosynthesis methyltransferase activity" EXACT []
synonym: "coenzyme Q biosynthetic process methyltransferase activity" EXACT []
synonym: "ubiquinone biosynthesis methyltransferase activity" BROAD []
synonym: "ubiquinone biosynthetic process methyltransferase activity" BROAD []
xref: EC:2.1.1.-
is_a: GO:0030580 ! quinone cofactor methyltransferase activity

[Term]

id: GO:0033460
name: GAG codon-amino acid adaptor activity
namespace: molecular_function
def: "A triplet codon-amino acid adaptor activity that recognizes a GAG codon." [GOC:mah]
comment: Note that in the standard genetic code, GAG codes for acid.
is_a: GO:0030533 ! triplet codon-amino acid adaptor activity

[Term]

id: GO:0043428

name: 2-heptaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity
namespace: molecular_function
def: "Catalysis of the reaction: 2-heptaprenyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-methionine = 2-heptaprenyl-3-methyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-homocysteine." [GOC:jl, PMID:11583838]
comment: Note that the polyprenyl sidechain substrate for these methyltransferases varies in length between species, for example, 6 units in *S. cerevisiae*, 8 units in *E. coli* and 10 units in *G. suboxidans*. Where the length of the substrate polyprenyl chain is unknown, the term '2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity ; GO:0008425' should be used.
synonym: "2-heptaprenyl-6-methoxy-1,4-benzoquinone methylase activity" EXACT []
xref: MetaCyc:RXN-9227
is_a: GO:0008425 ! 2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

[Term]

id: GO:0043429
name: 2-nonaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity
namespace: molecular_function
def: "Catalysis of the reaction: 2-nonaprenyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-methionine = 2-nonaprenyl-3-methyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-homocysteine." [GOC:jl, PMID:11583838]
comment: Note that the polyprenyl sidechain substrate for these methyltransferases varies in length between species, for example, 6 units in *S. cerevisiae*, 8 units in *E. coli* and 10 units in *G. suboxidans*. Where the length of the substrate polyprenyl chain is unknown, the term '2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity ; GO:0008425' should be used.
synonym: "2-nonaprenyl-6-methoxy-1,4-benzoquinone methylase activity" EXACT []
is_a: GO:0008425 ! 2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

[Term]

id: GO:0043430
name: 2-decaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity
namespace: molecular_function
def: "Catalysis of the reaction: 2-decaprenyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-methionine = 2-decaprenyl-3-methyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-homocysteine." [GOC:jl, PMID:11583838]
comment: Note that the polyprenyl sidechain substrate for these methyltransferases varies in length between species, for example, 6 units in *S. cerevisiae*, 8 units in *E. coli* and 10 units in *G. suboxidans*. Where the length of the substrate polyprenyl chain is unknown, the term '2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity ; GO:0008425' should be used.
synonym: "2-decaprenyl-6-methoxy-1,4-benzoquinone methylase activity" EXACT []
is_a: GO:0008425 ! 2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity

[Term]

id: GO:0043433
name: negative regulation of transcription factor activity
namespace: biological_process
def: "Any process that stops, prevents or reduces the frequency, rate or extent of the activity of a transcription factor, any factor involved in the initiation or regulation of transcription." [GOC:jl]
synonym: "down regulation of transcription factor activity" EXACT []

synonym: "down-regulation of transcription factor activity" EXACT []
synonym: "downregulation of transcription factor activity" EXACT []
synonym: "inhibition of transcription factor activity" NARROW []
is_a: GO:0043392 ! negative regulation of DNA binding
is_a: GO:0051090 ! regulation of transcription factor activity

[Term]

id: GO:0043434
name: response to peptide hormone stimulus
namespace: biological_process
def: "A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a peptide hormone stimulus. A peptide hormone is any of a class of peptides that are secreted into the blood stream and have endocrine functions in living animals." [PMID:11027914, PMID:15134857, Wikipedia:Peptide_hormone]
synonym: "response to polypeptide hormone stimulus" EXACT []
is_a: GO:0009725 ! response to hormone stimulus

[Term]

id: GO:0033937
name: 3-deoxy-2-octulosonidase activity
namespace: molecular_function
def: "Catalysis of the reaction: endohydrolysis of the beta-ketopyranosidic linkages of 3-deoxy-D-manno-2-octulosonate in capsular polysaccharides." [EC:3.2.1.124]
synonym: "2-keto-3-deoxyoctonate hydrolase activity" EXACT [EC:3.2.1.124]
synonym: "capsular-polysaccharide 3-deoxy-D-manno-2-octulosonohydrolase activity" EXACT [EC:3.2.1.124]
synonym: "octulofuranosylono hydrolase activity" EXACT [EC:3.2.1.124]
synonym: "octulopyranosylonohydrolase activity" EXACT [EC:3.2.1.124]
synonym: "octulosylono hydrolase activity" EXACT [EC:3.2.1.124]
xref: EC:3.2.1.124
xref: MetaCyc:3.2.1.124-RXN
is_a: GO:0004553 ! hydrolase activity, hydrolyzing O-glycosyl compounds

[Term]

id: GO:0046428
name: 1,4-dihydroxy-2-naphthoate octaprenyltransferase activity
namespace: molecular_function
def: "Catalysis of the reaction: 1,4-dihydroxy-2-naphthoate + polyprenylpyrophosphate = dimethylmenaquinone + pyrophosphate + CO2." [MetaCyc:DMK-RXN]
synonym: "1,4-Dihydroxy-2-naphthoate prenyltransferase activity" EXACT []
xref: EC:2.5.1.-
xref: MetaCyc:DMK-RXN
is_a: GO:0002094 ! polyprenyltransferase activity

[Term]

id: GO:0050534
name: 3-deoxyoctulosonase activity
namespace: molecular_function
def: "Catalysis of the reaction: 3-deoxyoctulosonyl-lipopolysaccharide + H2O = 3-deoxyoctulosonic acid + lipopolysaccharide." [EC:3.2.1.144, MetaCyc:3.2.1.144-RXN]
synonym: "3-deoxyoctulosonyl-lipopolysaccharide hydrolase activity" EXACT [EC:3.2.1.144]
synonym: "alpha-Kdo-ase activity" EXACT [EC:3.2.1.144]
xref: EC:3.2.1.144
xref: MetaCyc:3.2.1.144-RXN
is_a: GO:0004553 ! hydrolase activity, hydrolyzing O-glycosyl compounds

[Term]
id: GO:0004673
name: protein histidine kinase activity
namespace: molecular_function
alt_id: GO:0008896
def: "Catalysis of the reaction: ATP + protein L-histidine = ADP + protein phospho-L-histidine." [EC:2.7.13.3, GOC:mah]
subset: gosubset_prok
synonym: "ATP:protein-L-histidine N-phosphotransferase activity" EXACT [EC:2.7.13.3]
synonym: "EnvZ" NARROW [EC:2.7.13.3]
synonym: "histidine kinase (ambiguous)" EXACT [EC:2.7.13.3]
synonym: "histidine kinase activity" EXACT []
synonym: "histidine protein kinase (ambiguous)" EXACT [EC:2.7.13.3]
synonym: "histidine protein kinase activity" EXACT [EC:2.7.13.3]
synonym: "HK1" RELATED [EC:2.7.13.3]
synonym: "HP165" RELATED [EC:2.7.13.3]
synonym: "protein histidine kinase (ambiguous)" EXACT [EC:2.7.13.3]
synonym: "protein kinase (histidine)" EXACT [EC:2.7.13.3]
synonym: "protein kinase (histidine) (ambiguous)" EXACT [EC:2.7.13.3]
synonym: "protein-histidine kinase activity" EXACT []
synonym: "Slnlp" NARROW [EC:2.7.13.3]
xref: EC:2.7.13.3
xref: MetaCyc:2.7.13.3-RXN
is_a: GO:0004672 ! protein kinase activity
is_a: GO:0016775 ! phosphotransferase activity, nitrogenous group as acceptor

[Typedef]
id: negatively_regulates
name: negatively_regulates
is_a: regulates ! regulates

[Typedef]
id: part_of
name: part_of
xref: OBO_REL:part_of
is_transitive: true

[Typedef]
id: positively_regulates
name: positively_regulates
is_a: regulates ! regulates

[Typedef]
id: regulates
name: regulates
transitive_over: part_of ! part_of